Background Notes on The Literature of Trust: Bridging The Perspectives of Socio-Economics and Technology

Mark Burgess

March 13, 2023

Abstract

This document is background material for the forthcoming Promise Theory of trust. It contains notes on the research literature from a number of disciplines. It's secondary motivation is to comment informally on the extent to which independent research is compatible *a priori* with Promise Theory, it's definitions and its predictions. A precise summary in Promise Theory is work in progress and will be reported elsewhere.

The supposition that trust is a purely human phenomenon is ubiquitous outside Computer Science, and undermines the generality of its significance for human-technological interactions; thus, it is of interest to abstract away specifically human attributes and identify the signalling mechanisms that generalize trust for 'agents' in the general case.

Trust emerges as a way of trading 'savings of effort' (process debt) for possible gain. As such it's an important cost saving strategy for agents that have finite resources, and which therefore need to prioritize activities of greater value. This is a way to avoid being taxed by accountability and verification protocols to hedge against uncertainty. The story is linked to the 'tragedy of commons' or shared resource depletion, in this case where a single agent's resources are being shared between possibly competing tasks.

In terms of process models, the appearance of an action potential that we call 'trust' is a sign that we are dealing with learning processes, also called *memory processes*, not transactional Markov processes.

In order to understand how we shall use trust to lubricate human-technology (cyborg) relations in an increasingly augmented semi-virtual world, we need a consistent and formalized model of trust that everyone could agree on. In particular, we look for a way to avoid using moral judgements in trust questions as these are frequently misleading.

Contents

1 Introduction

This document is not a scientific paper or a set of lecture notes; rather, it is a sketchy set of personal notes with commentary about the state of research literature about trust, spanning a number of disciplines. They are open and available for the benefit of others who might find them valuable.

The notes have a secondary agenda: to try to assess the extent to which Promise Theory's model of trust is supported (or not) by the multitude of views and experimental findings. The result is that the Promise Theory model is well aligned with the breadth of the literature. The notes will serve as

supporting background for a separate series of documents about the Promise Theory of Trust, to be given elsewhere.

The subject of trust has a large and dispersed literature and is fairly recent, being influenced by two key books written in the 1990s, by Fukuyama and Putnam respectively. Later work in business literature and computer security literature took on the question in relation to its own challenges, without reference to these studies.

1.1 Social origins

In sociology writings, trust is part of a larger debate about Social Capital. The definitions of these terms are hard to pin down, but a discussion of the latter is found in [1, 2]. One has the impression that sociologists are interested in documenting the many icings on a cake rather than the recipe of cake itself. This is a natural beginning, as in zoology where one tends to start with listing species to document what one can.

My feeling is that we can do better, drawing inspiration from Newton [3]. The broadest descriptions of social capital go beyond the role of trust as a lubricant of interactions. Empirical studies in social science tend to be based on questionnaires that pry self-assessments from willing participants. This form of data collection is coarse but it's compatible with the notion of assessment in Promise Theory. The absence of independently observed measurement data is a weakness that may make the search for a common theoretical model less appealing to some. Nevertheless, it remains compelling to me, and Promise Theory seems to offer a very helpful structure.

1.2 Information Technology

In Computer Science (CS) and information technology (IT), trust is quickly swept under the cryptographic rug, by reducing assessments of trustworthiness to a crypto token or signature that confirms some rudimentary validation of identity. In other words, a trust token is a receipt for minimal effort in validating some agent's identity credentials. Once identity is confirmed, an agent is assumed sufficiently trustworthy to pass through a gateway. Access controls are then used based on the confirmed identity to extend trust (or not) to those who pass this basic test. Thus, the idea is that trust acts as a reward or till receipt for some process for due diligence. Clearly, the meaning of trust, in a societal sense goes far beyond such a token [4]: being relatively certain about someone's formal identity says nothing about how they are going to behave. The distinction is only "good guy / bad guy" as a simplistic Hollywood caricature of risk. There is a hope that trust encourages cooperation and avoidance of attack. If we are to develop a deeper understanding of trust in a cyber-networked society, a much more diligent view of trust is needed. The aim of these notes is to synthesize a minimally acceptable meaning for trust, so that it can be applied to the dynamical questions about life online.

1.3 Game theory

The history of rational approaches to trust has become closely associated with the Prisoner's Dilemma model in Game Theory [5,6]. This is a natural angle, bolstered by one of the few analytical and modelling frameworks available from the days of cold war politics [7, 8]. Today, Promise Theory also provides an additional layer underpinning games [3,9,10], which can deepen the game theoretical result. The narrative is extended to study the ins and outs of contracts, including optimizations of negotiations with moral hazards and principal agents [11, 12].

Many authors treat Game Theory with a particular authority, probably more due to the cachet of its progenitors than the utility of its findings. Game Theory supplies a partial framework for understanding benefit and loss in synchronous and asynchronous interactions, by incorporating private economic value into decision to align or deviate from an abstract course of action (this is mainly in relation to the ubiquitous Prisoner's Dilemma model, which is the Simple Harmonic Oscillator of socio-economics), but it has basically nothing to say about context or semantics, which are so richly pursued in the social sciences, for instance.

A more sociologically sensitive use of this kind of economic thinking was examined in the writings on Elinor Ostrom, Nobel Laureate, concerning the 'tragedy of commons', i.e. the sharing of common resources, which again is closely related to mutual trust, once again with a game theoretic framing [13,14].

1.4 Memory processes versus transactional Markov models

In Hamilton and Axelrod's versions of Game Theory for selection [7,8], the response to a breach of trust could be a tit for tat strategy. This is directly related to machine learning techniques. Game theory avoids validation or verification and learns directly from the Markov process of outcomes from recent game plays. This kinds of Markov retaliation is cheap to remember, so it saves the agents from knowing the identities of a lot of individuals in terms of processes. The process of reason means that precautions have a price. The mechanism that connects interpersonal trust with engagement and repeated interaction has its roots in the Prisoner's Dilemma model. We know, from Dunbar's studies, that humans have finite resources to remember and build trust in others. The Dunbar hierarchy sets an approximate limit on the number of relationships a person can maintain knowledge about (at different levels of detail) [15, 16].

Trust as a stochastic process [17] suggests we might not need memory to know too many individuals, as long as we accept scaling limits, tribe sizes, group cooperation limits, etc.

1.5 Is trust a universal scale of measurement?

Most authors seek models of trust that appear universal and homogeneous for all agents. This is what one would expect of elementary Markov processes, but more complex processes in the natural world are memory processes, and these are explicitly path or context dependent. For trust, there are strong indications that a universal memoryless model is inappropriate.

At the scale where trust plays a role, systems are far from universal or impartial, but there may still be deeply underlying dependencies at work, shaping the dynamics. Game Theory models, like Prisoner's Dilemma do not require the payoffs to be delivered in a common currency. Each agent maintains its own ledger of play, so there is no conflict there.

The resolution of such issues are what Promise Theory may help to uncover. For instance, some seek to rely on a notion of reputation as a global authority, assuming that this must be impartial, but evidence from studies as well as theory suggest that trust is a purely individual agent-based assignment, based on contextual parameters and that private. Reputations too are subject to local resonances through superagency, based on the intermediate concept of authority [18] (force and power) and circumstance, which invalidates the assumption of homogeneity and thus impartiality at the level of humans.

There are two issues that undermine homogeneity or impartiality:

- Acceptance of advice from recommendations, reputation, and other sources, which may be subject to spurious and unstable resonance effects, and
- The fundamentality of local trustworthiness assessments, which is the ground stats behaviour of all agents.

1.6 Power, authority, and resonant interactions

The concept of 'power' in a social or political sense is related to trust. The rise of powerful figures depends on either

- The ability to dominate (impose)
- The willingness to accept imposition (conscious or unconscious), e.g. hopefulness
- The intentional promise to align with a leader as a follower.

As certain promises are kept, more followers will join in. The tendency to trust based on referred assessment (reputation, mass suggestion, etc) will create resonant levels of trust that strengthen or suddenly weaken a powerful agent. Sometimes we trust powerful figures, sometimes we mistrust them. Power implies trust but not necessarily capability.

An assessing agent can attempt to distinguish power over things that benefit me (hope), versus power over me (risk).

Social power may be resonant, if it depends on support from a group. A leader is supported by followers, and therefore has power over them and on behalf of them. That make a double edged sword, so the interaction may resonate to build trust or to destroy trust.

Power is related to authority, i.e. the holder of support in the form of bulk promises to accept impositions from the authoritative agent [18].

1.7 Currency exchange in trust

The issue of how different trust potentials, for different interactions, mix and merge is an important topic. The fact that agents may exchange trust in one thing for trust in another, by making some kind of equivalence in their assessment algorithm, allows trust to be 'hacked'. Marketing, propaganda, phishing, etc all use this vulnerability to induce cooperation.

In economic terms, trading trust is not like trading common money, it's more like exchanging different currencies between different economies, because each agent has its own private valuations in interactions, and there is no external bulk market in which different currencies might equilibrate to a consensus on how much one kind of trust is worth compared to another.

Agents may use on currency (or form of social capital) to pay for another. Thus, trust in a group might be exchanged for trust in a leader's own qualities. This allows resonances to 'hack' the rationality of an assessment. In other words, an agent might feel happy one day and simply go along with a questionable decision, because it is in a good mood. Assessments merge and mix together within an agent, depending on its internal processes.

1.8 Socio-technical trust

In the modern world, there is a need to scale trust in large numbers of interactions through technology and its automation systems, because human to human interactions are either impractical, too slow, or too costly to scale in mass production. This difficulty can't be solved by employing more people in systems because there may not be sufficient numbers of trained individuals, and we seek consistency in interactions in order to win trust. Humans are naturally inconsistent for many reasons.

The purpose of these notes is to review and capture the relevant aspects of trust literature that relate to how trust can be assessed and measured, both by individuals involved in a process and by non-partisan experimenters without a stake in the outcome.

2 Promise Theory Background

Promise Theory is a convenient and flexible semi-formal language with which to describe a theory of trust. In this section, we pick out some key themes that can be represented in Promise Theory [10]. Promise Theory doesn't claim to formalize trust uniquely. In fact, it predicts that trust is a uniquely individual policy, based on private assessments, shared information, and ultimately individual intentions and goals. Rather, it provides a consistent framework for building the common elements of different kinds of trust so that agents can be understood in context based on common principles. Those principles have to do with prioritization, economic activity costs, and alignment of preference semantics. There is a tradeoff between trust and cost of control [19].

2.1 Promise Theory definitions, hypotheses, and predictions

There have been few serious attempts to mathematize trust. Normally, authors will defer to probabilistic arguments, but one rarely sees how probabilities are defined. We can do better using Promise Theory. We begin by defining a promise between two agents S (Sender or promiser) and R (Recipient or promisee).

- A promise is the measuring stick by which we assign meaning to outcomes. In a 'donor' promise from S to R,

$$\pi^{(+)} : S \xrightarrow{+b_{\text{offer}}} R, \tag{1}$$

the body b_{offer} is accepted at a level of its own choosing, from the maximal content of the set, with a 'receptor' promise:

$$\pi^{(-)} : R \xrightarrow{+b_{\text{acceptance}}} S, \tag{2}$$

so that the promised interaction has the strength

$$b(S, R) = b_{\text{offer}} \cap b_{\text{acceptance}}. \tag{3}$$

Over a number of reliances on this binding, any agent in scope can sample or assess the extent to which the promises were kept. This is written $\alpha(\pi)$. Each agent A will form its own private assessment, which one could associate (as one example) as the dimensionless frequency or probability:

$$\alpha_A(\pi) \overset{e.g.}{\mapsto} \frac{\text{samples where promise kept}}{\text{total samples}} \tag{4}$$

- Trust is related to this assessment of a promise being kept as an expectation informed by this assessment. If we think of the saying "trust but verify" and "if you don't trust you have to verify, and verifying is expensive", we're motivated to define the sampling rate for verification:

  $$\text{Sampling rate for } \pi \propto R_0(1 - \alpha(\pi)) \tag{5}$$

  as an inverse level of activity invested by R, or assigned by R to interactions with S, based its the expectation of the promise being kept. The maximum value R_0 is based on the interior resources of the agent. It cannot be constant or homogeneous between agents, as it's an internally shared resource. In this way, trust becomes a mitigating factor to excuse verification. There are various ways in which it could be defined, but this linear approach makes reasoning simpler. The more R trusts S, the lower the sampling rate for verifying the promise to be kept. The less an agent trusts another, the higher the sampling activity or 'energy' level of the agent in sampling $b(S, R)$. It implies that trust is related to a process, which is a step forward from intrinsic static properties.

- Trust in either self or others may attributed to specific entities or their promises of processes. Ultimately trust in process is the fundamental idea; trust in entities is the trust is the assessment of an entity concerning some specifically promised process.

- We can't trust someone when they make promises that violate the first law of promises, i.e. that agents cannot make a promise on behalf of anyone but themselves. e.g. I promise that you won't get hacked. I promise that you are safe.

  These are promises about someone's assessments at best. Maybe just white lies.

- Trusting is *kinetic* because it requires work on the part of the assessor R to maintain a sampling of outcomes over time.

  Trusting involves the placement of a bet on trustworthiness, a willingness to invest in another agent. A trusting agent assesses whether promises are kept less frequently, i.e. has a lower Nyquist sampling frequency. Thus trustworthiness is the complementary assessment, like the return on the investment of trust, except that it is the extent to which the investment of 'absence of work' to verify an outcome was justified.

- Trustworthiness and risk are assessments of the estimated integrity in keeping promises.

Trust is a policy expressed by processes between self and non-self agents. It may involve internal resources to assess or implement. As usual, the concepts of space and time play the fundamental roles in processes.

- Space: a trust boundary is a group boundary, or a lateral trust boundary, formed by membership promises.

- Time: a trust horizon is a timelike or longitudinal boundary on trusting behaviour over iterations of a relationship, such as Axelrod game rounds.

Trusting agents have a longer spatial or temporal horizon than untrusting agents.

Remark 1 (Trust Horizon) *Define the trust horizon for an agent to be an intrinsic promise of the maximum number of interactions before giving up on a relationship.*

The trust horizon might affect the trustworthiness of agents to others too, since giving up on a relationship might alter the willingness of the agent to accept or keep promises in the future.

The trust horizon does not imply transitivity of trust. Note the end-to-end problem discussion in [10].

2.2 Reputation, confidence, and belief in outcome, hope

In Promise Theory, trust acts as an underlying "turtle" on which many others, i.e. definitions rest. The original promise trust paper [9] outlines some of these.

The idea of a reputation suggests there is a way to measure agents' intrinsic trustworthiness by using an ensemble of agents to form assessments, which are then aggregated and passed on somehow. Bulk measurement is a classical strategy to normalized repeatability, but this normally applies to repeatable conditions. Agents may not be elementary and homogeneous as in physics experiments, so it might not be easy to measure a fair value for trustworthiness. Each assessment may have different semantics.

We know that reputation can be made self-consistent [9], in the sense of an eigenvalue problem, but this is at the expense of the assumption of equivalence between agents. Many authors treat the idea of reputation as significant, probably because they accept that it's ad hoc. The principle of autonomy implies that reputation should not be considered to be mandatory for an agent; it is only advisory—there can be no compulsion to accept trustworthiness. The final decision is made autonomously. Agents may be willing to engage until they judge that the relationship has no value to them, where value may be measured in money, in kind, etc.

Remark 2 (Trust) *Trust relates to the assessment that a promise (no matter how absurd) will be kept.*

For example, Foxx will be consistently opinionated, Reuters will be be consistently close to neutral.

There is a subtle difference between having confidence in someone's ability to keep a promise (expectation) and trust. Trust motivates expectation but is not a complete assessment of capability (or competence). In the future, as we try to automate trust using AI, the training biases of AI will make this selection of confidences even less transparent. Information is part of a supply chain, in which each step leads to a game of 'Chinese whispers'. This is the intermediate agent theorem.

We can assume that trust underpins confidence, since without trust one would not assess confidence in an outcome. Confidence is "trust AND capability".

Remark 3 (Confidence) *Confidence relates to the assessment of an agent's ability to promise competence or ability perhaps conditionally in a particular context.*

Thus we can consistently formulate this as a logical (and thus probabilistic) construction, where confidence is an expectation value using trust as a probability of a specific promise being kept. Confidence excludes an assessment of the credibility one's own assessment(!) So we might add to that "belief in outcome" as

$$\textit{"trust in self"}\ \textbf{AND}\ \textit{"trust in other"}\ \textbf{AND}\ \textit{"capability of other"}. \tag{6}$$

Then using the assessment of promise keeping for π as the measure, belief becomes:

$$= \alpha_{\pi^{(-)}} \times \alpha_{\pi^{(+)}} \times \alpha_{\pi} \tag{7}$$

For example, in the modern world we have to assess whether news is fact or propaganda. We might trust the news to be good or bad (say Reuters vs Foxx News) but we also need to assess our confidence in the promised material.

Remark 4 (Hope) *Some authors also use trust in a hopeful sense, when they suggest that trust must come from a position of vulnerability. This seems to confuse trust with hopefulness, which can be accounted for separately and logically if we postulate a measure of well-being to assign semantics.*

Trust might be artificially amplified by the reduction of wellbeing; e.g. something like this:

$$\textit{Hope} = \frac{\textit{confidence}}{\textit{well being}} = \frac{\textit{other-confidence}}{\textit{self-confidence}} \tag{8}$$

The scales here have to be defined for a given agent.

For example, the more critical a resource is (water, electricity, food supply, "basic human rights"), the more we are inclined to trust it and feel victimized if the promise of it is not kept. This is a key challenge to society, as trust leads to passivity. This leads to the strategy of 'imposing expectations' or (-) promise assessments as a weaponization of trust.

Hypothesis 1 (Trust and passivity) *The expectation of a good or service being delivered may lead to passivity and thus increased trust or risk appetite, contrary to the Downstream Principle in Promise Theory, which suggests that agents should work to receive. Thus trust undermines the downstream principle, which in turn leads to impositions of blame and accusation.*

2.3 Impositions reduce trust

1. Misalignment between impositions and receptors tends to reduce trust for donor in recipient.
2. Group promises are easier to secure when they are aggregated with a coarse receptor (broad base of support).
3. Singular promises offered to a broad base of agents (coarse group) will receive more support than
4. Resonant relationships can align dynamically starting with approximate or weak alignment, if promises are initially kept.

   Note the findings in [20]: in organizations, judgements are often made by groups. There are resonance phenomena within the group, not just between leader and follower, but forming an echo chamber around "us versus them" issues that are not intra-group but inter-group dynamics.
5. Trust takes longer to establish than to revoke, because it builds from weak coupling on average so commands less attention by an agent, in order to introduce new promises in a relationship.
6. Trust can be revoked quickly by a single agent. There will be some inertia for trust to be grown or revoked in a group setting, so the larger the group, the longer the tail for trust to be dissipated.

   Note that it's not trustworthiness (potential trust) that's dissipated initially, but kinetic trust or risk appetite. Potential trust is a local property assessment, seen through the lens or filter of a binding interaction.

We can note the role of accusations in trust dynamics too. An accusation is a special kind of imposition in general. It might undermine the trustworthiness of the accuser, whereas a promised accusation might be considered more trustworthy.

2.4 Potential and kinetic trust, trustworthiness as a potential function

The appearance of an action potential trust is a sign that we are dealing with learning or *memory processes*, not transactional Markov processes. The memory of a process can be within an agent, or without, i.e. between agents. Through the promises they make, agents serve as memory for one another—as part of their extended state, as in any physical system.

We recall that trust is an assessment, or book keeping value, not a resource as such. Indeed, it replaces activity in that the penalty for not trusting is to require action, which does rely on actual (and generally finite) resources of an agent downstream of a promise. From [3], we can summarize:

- Kinetic trust: Promising the assessment that an agent is probably trustworthy is the same as promising kinetic trust. This is equivalent, by some transformation, to risk taking.
- Potential trust (trustworthiness) is an assessment by a downstream receptor agent, with a (-) promise, of an upstream agent's ability to advance a process in order to keep a (+) promise. By complementarity $(upstream, -) \leftrightarrow (downstream, +)$.

Hypothesis 2 (Trust in Low Information) *The less specific the promises made by an agent, the cheaper and easier it is to assess it as "probably trustworthy", since there is less information and thus fewer constraints to measure it against. Conversely, agents that make very specific promises are harder to trust.*

Thus there is a tendency to trust compressed information that one can "wrap one's head around" or afford to process.

A corollary of this is that the more internal semantic diversity a group can promise, i.e. the more entropy or process complexity in its promises, the greater the cost in assessing in detail, thus there will be tendency to coarse grain and approximate in order to assess and trust, especially in an agent with limited resources. Conversely, agents that promise nothing specific or promise a diverse portfolio which is costly to assess are likely to win trust without evidence. This is like "blinding with science".

Human agents judge with their emotions, if we accept Mercier and Sperber's evidence. The Semantic Spacetime hypotheses also align with this. Thus agents will tend to downgrade rational assessment when it is costly. System 2 gives was to system 1.

2.5 Group dynamics, stereotyping, and tribes

We would like to know how trust scales in groups. Promise Theory has a hypothesis about this, but it is not applicable to the group characterizations in the literature, which are more about a supposition of identity politics (why certain groups don't like each other).

Very little in the way of group dynamics seems to be available from sociological studies. In biology, we have studies of animal group behaviours that could, in principle, be addressed by currency arguments [21]. It's a matter of definition whether one believes animals can exhibit trust. There's a history of differentiating human behaviours from the rest of the animal kingdom, but this seems dubious.

In [21], the simulation modellers found that "if individuals pay attention to the actions of their fellow group members, they will tend to remain at a safe location for longer. This demonstrates that simple social behaviours can result in the repression of consistent inter-individual differences in behaviour, giving the first theoretical consideration of the social mechanisms behind personality suppression". If we couch this in terms of the trust dynamics of [3], then it suggests that agents who promise to align with local promises, i.e. with group receptors, may have a survival advantage. So such trusting behaviours self-select for survival value. Risk averseness is a group behaviour, while risk taking is an individual pursuit, so if an individual (say alpha) takes a risk to stand out, others may follow on average—which offers a mechanism for leader selection. The 'leadership mandate' is simply the willingness or propensity to listen to others.

Promise Theory thus predicts that non-group members cost more to evaluate, as they can't be reduced to a group assessment, so exterior arrivals might tend to grab more attention than the equilibrated masses in a group, and become naturally predisposed to becoming leaders, e.g. the appearance of an alpha male, or an intruder. Leadership characteristics will tend to emerge if the main population has receptor promises for an attribute that an individual can impose upon. The receptors give the mandate of authority, in the sense of [18].

However, there are competing forces: if the entropy hypothesis is true, agents will tend to trust the group more than the individual, so a leader has to display group qualities or exceptional return on the investment of taking the time to judge. The same is true in job interviews.

- Reputation systems as well as trust acceptance checkpoints are authoritarian overrides of trust. Please sign away your own opinion by subscribing to mine, or tell us the consensus of opinion by the same rules.
- Peer recognition, peer pressure.
- Group identity Domination in nates.
- Leader selection, emergence of authority.

2.6 Diversity in groups, institutions, and coarse grained entities is largely unexplored

It's common to associate agency with abstract groups of agents related to a process. For example, we say "the market reacts" to changes, as the invisible hand of economic activity. This is reasonable, as the collective group of agents forms a larger inter-agent process, with a definable if fluid boundary. Thus we can view the collective process as an internal process of a superagent instead.

How we deal with coarse grained entities (groups, institutions, nations, etc) is an important element of trust, which seems hardly discussed in the literature. To some extent there is a discussion of identity groups, such a racial minorities based on the assumption that these are the relevant or appropriate distinctions for trust. That seems like a missed opportunity, given the scaling predictions of Promise Theory [10].

Much of the research in social science, ignores the effect of process, and looks only to identity politics of groups. Sociologists seem interested only in the politics of ethnic identities rather than collaborative processes like companies [22]. This identification places a shared identity group in a similar frame as a named individual, i.e. the group identity is a unit of trust. The principle ought to extend to any group that interacts through a network of promises as long as we can define the edge.

The scaling law of group activity is that shared processes tend to accelerate in groups due to parallelism and lack of semaphores to limit progress. This will affect sampling rates and assessments, and thus trust too. Similarly, trust in something less specific feels easier to grant as it has fewer obvious features to attach semantics to and thus criticize rationally. These questions need to be considered and verified. The literature so far only hints at certain answers.

In technology, the coarse shared entities include the certificate authorities. While they can be imperfect and even corrupted, they generally play a positive role in enabling economic activity. Being somewhat invisible and non-descript, they are easy to trust. Their broad support from customers supports the idea that they are trustworthy by vote—just as influencers and leaders with large numbers of followers seem to be more trustworthy due to the large number of unspecified promise relationships. It might be a cognitive deficiency (approximation) which leads to trusting based on number. That's surely something that could be tested in experiments.

- Participation in groups is a key topic in social trust. Participants' choices were assessed to be cooperative, competitive or individualist, risk seeking, assessing their risk appetite. To what extent can these measures be operationalized, i.e. turned into concrete judgements or outcomes? e.g. in political voting or signing up for a membership or participation in some happening. Social scientists typically decompose political participation into three components: voting, institutionalized participation (e.g., contacting politicians, working for political campaigns), and non-institutionalized participation (e.g., signing a petition, joining a boycott).
- Although far from conclusive, social science researchers generally find that generalized trust (i.e. a positive attitude towards people in general) is positively related to non-institutionalized participation (e.g., Kaase 1999) and either negatively related or statistically unrelated to institutionalized participation (e.g., Back and Christensen 2016). This initially suggests a result which tends to not support the coarse trust hypothesis. However, if we view institutions as agents with personalities in a particular framing, then it still makes sense. This tells us that it isn't the size that matters as much as the entropy or lack of identity.
- Cultural baselines in attitude would play an important role to the semantics of trust. In Germanic countries, the 'culture' of unwillingness to pay for a partner, but rather to split a bill and pay only for oneself, might seem to indicate a low trust. On the other hand there is a high degree of trust in formal institutions in Germanic countries. In Asia and Middle East, the opposite is perhaps true. Generosity is a cultural expectation, indicating honourable behaviour. On the other hand, trust in government is marred by corruption. The culture of generosity could be seen as a factor in corruption, i.e. the expectation of gifts.

One should be cautious in attributing causality to group identities as there is a tendency to do in social studies. It could be that the size of groups is the factor rather than their makeup. This is hard to assess from the literature, since the assumption is that the characteristic semantic labels are the causal factor.

Using behaviour to measure trust is not without its difficulties, as perception of intent is not the same as intent. In a Promise Theory sense, how information is received or acted on is purely a matter for the receiver (The Downstream Principle [10]). Measures of past behaviour are considered more trustworthy(!) than speculative attitudes.

The apparent tendency to scale trust with entropy, i.e. to offer more kinetic trust or assess greater trustworthiness to diverse or non-descript entities appears when voting or agreeing with uncertainty. "Someone will buy your house" is usually assessed as more trustworthy a promise than "Martin will buy your house". This profiling or patterning applies in units of separate assessments, not in terms of the bulk size of a thing. Thus, we treat an individual of large size with equal weight as one with small size, since it's the cost of assessment that matters for trust. Clearly one may does not trust an elephant more than a mosquito on the basis of size alone, but one trusts mosquitos in general more than one trusts a single mosquito to promise what mosquitos promise.

Hypothesis 3 (Entropy and trust neutrality) *Constant churn of associations, forming and reforming, as well as loss of process memory, all lead to widespread resetting of trust to its default value, i.e. to that initial policy that applies for each assessor.*

While uniqueness may catch the attention of an agent and resonate with its attention, uniqueness generally seems to be a negative attribute for trust, viewed relative to a more normalized promise. Thus, we might expect uniqueness to play a role in resonant exploratory risk taking, but ultimately be downgraded when pitted against norms ("tried and trusted" but not necessarily by me).

3 Trust in social science

Two books published in the 1990s have been particularly influential in social science trust studies. These are *Making Democracy Work*, by Putnam [23] and *Trust*, by Fukuyama [24]. They are widely referred to

in sociological literature, but not at all in Computer Science. These books have been compared in review in [25]. This review borrows from some earlier reviews in an effort to be concise.

In the view of social sciences, trust is widely assumed to arise when a community shares a set of moral values so as to create the expectation of regular, reliable, and honest behaviour (promise keeping). Fukuyama calls familial dominated societies "low trust" societies, characterized by short range trust. Long range trust coherence can only come from voluntary cooperation, in his mind. This is a political bias, however, as Promise Theory also allows a solution based on imposition if the general population is sufficiently pliable in terms of receptor promises. This is the case in feudalism, for example.

The concepts associated with trust are often coloured with preconceived social and political ideas, such as civic duty and government, unlike the idea of a generic kind of trust that motivates and shapes behaviour on a deterministic level as discussed in [3].

From [22], we get some insight from a sociologist about the conflicting assumptions about trust. It's worth noting the immediate identification of trust with the concept of voluntary cooperation, which is compatible with the lens of Promise Theory [10].

> Putnam (1993, 1995) popularized the idea that people learn trust from participating in voluntary organizations. A robust civil society may therefore be one factor which creates generalized trust, although evidence on this point is mixed (Nannestad, 2008). Other researchers found that generalized trust tends to thrive in conditions of low income inequality (Bjørnskov, 2007; Kawachi et al., 1997), and where the rule of law is strong (Knight, 2001), while others have identified an association between high trust and Protestant religious traditions, supposedly because of Protestantism's emphasis on equality and accountability to God (Delhey and Newton, 2005). The determinants of generalized trust may also differ between Western and non-Western countries (Freitag, 2003). Many of these studies fall into the experiential' school of trust scholarship, which highlights the role of life experiences in shaping individuals' generalized trust (Wu, 2020). But other scholars disagree that trust is the product of one's experiences at all. Uslaner (2002) contends that trust is best understood as a personality trait that one is either born with or which crystallizes early in childhood, and does not change much in response to particular experiences. There is some evidence in favour of this view, such as Dawson (2019) and Stolle and Hooghe (2004), the latter of which notes that generalized trust shows a large degree of stability' over time. This view of trust has become known as the cultural' theory

This excerpt suggests that personal biases may influence experiments in social context–a case of seeing what one wants to see. In particular, Fukuyama writes from a strongly American perspective that doesn't always feed more broadly credible.

Trust is clearly tied to social groupings, and is challenged by racism, tribalism, bigotry etc, but the direction of cause and specifics of correlation are unclear. If we want to treat social capital, or trust in the broadest sense, as a basis for behaviour, then we need to look more deeply than these superficial social interactions.

3.1 Assumptions, definitions and measurement

Trust is essentially qualitative rather than quantitative in social science, mainly due to the difficulty of measuring such an intangible. One is always free to convert a quality into a quantity ad hoc, and this is common practice for statistical analysis. Trust plays a role in deciding and interpreting experiments too, including the assessment of trust itself. This leads to some interesting questions about self-consistency of interpretation.

The concept of risk is used indirectly in many social capital papers to discuss the influence of risks, such as burglary and violence, to attitudes to trust. Trust is closely associated with group dynamics and group identities. Discussions of scaling trust to groups seem to be exclusively related to the identity of groups, such as "shall we trust one legged people" rather than can be trust a certain number of people. This reflects the biases of social science and misses out on important questions about dynamics.

The concept of a trust radius captures the idea that trust may be limited to a particular relational distance within some group boundary, e.g. friend of a friend. There are many hopeful arguments for the transitivity of trust, which are quite unconvincing. Promise Theory predicts that trust is not transitive in any meaning sense, though there could be sporadic examples where it appears to extend transitively due to chance or external constraints.

Measurement of trust empirically is linked with Social Capital in general, where concepts of trust are in abundance and are freely mixed with a more general idea of social capital as a basis for power. The idea is to ask participants to self-assess trust and report on a graded semantic scale: yes, no, partially, etc.

Trust seems to be classified with a few coarseness levels:

- General trust in others (people)
- Trust in individuals.
- Trust in government, etc.

Most attempts to measure trust do so by questionnaire and the trust of an honest reply! The normalized model for most trust studies simply asked people for a poll of opinion: Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"

Subjectivity and individuality in responding and hearing the responses makes the notion of constant experimental conditions difficult to define or trust i social sciences, but this is just par for the course Typical questions are like this:

- How often do you lend money or personal possessions to your friends?
- How often do you leave your home intentionally unlocked?
- Would you give a password or a credit card to a friend or associate?

To relate these questions to trust, there has to be a model of the the thought processes involved: an algorithm or formula for computing. There is no simple linear relationship. This may be somewhat ad hoc.

In the social science literature, a lot of weight is given to two books: Putnam [23] and Fukuyama [24].

3.2 Putnam, Making Democracy Work, 1993

Putnam studies the rise of democracy in post war Italy, using the framework of Prisoner's Dilemma as a model. This game theoretic framing was especially popular from the 1970s owing to work by Axelrod on cold war dynamics [7]. Putnam's argument is that social capital is a necessary ingredient for government functioning [23]. Causality is cyclicly reinforcing. From the conclusions:

- In all societies, dilemmas of collective action hamper attempts to cooperate for mutual benefit, whether in politics or in economics.
- Third-party enforcement is an inadequate solution to this problem. Voluntary cooperation (like rotating credit associations) depend on social capital. Norms of generalized reciprocity [for favours received] and networks of civic engagement encourage social trust and cooperation because the reduce incentives to defect, reduce uncertainty, and provide models for future cooperation.
- Putnam considers trust itself as an emergent property of the social system, as much as a personal attribute. Individuals are able to be trusting (and not merely gullible) because of the social norms and networks within which their actions are embedded.
- Putnam laments the issue of measurement: since trust is so central to the theory of social capital, it would be desirable to have strong behavioral indicators of trends in social trust or misanthropy. I have discovered no such behavioral measures."

The differing patterns of social capital in the North and South are due to centuries of history, yet improved institutions have a gradual effect on improving social capital.

He views social capital as simply one of two equilibria: either societies choose 'always defect' in their daily collective action problems, or they choose 'always return favours', thus building social capital and general trust. Keep in mind: like all equilibria, these are self-reinforcing. That means that saying institutions cause social capital which reinforces institutions isn't necessarily circular; any equilibrium is circular in that sense, since being in the equilibrium increases the probability that you will stay there.

Putnam's innovation was to use social capital, through the lens, of game theory as guide to economic activity. This allows a simple Promise Theory framing.

3.3 Fukuyama, Trust, 1995

Valadbigi and B. Haratyunyan review Fukuyama and make a central observation that the scaling of trust has passed through a number of transitions: from small familial and kin relationships to the favour of impartial institutions of government that tend to wipe out such forces as "corruption", to a final stage of voluntary private institutions or corporations. They argue that he considers economic success to depend on the setting aside of familial values for the ultimate formation of purely voluntary associations, through impartial private economic entities, i.e. corporations.

Fukuyama's writings often seem to be a little too supportive of American ideologies to be fully neutral in their assessments. Another example of this was his earlier book, *The End of History*, which basically said that capitalism won over socialism. His answers are conveniently aligned with the idea that the American system is the right one, to which all others will flow. The criticisms of socialism, in the sense of Eastern Europe and Soviet societies, describe only the destruction of civil society as interpersonal associations were displaced by relationships with institutions, which hampered market economies. Their replacement by official public institutions (which are implicitly negative) is viewed as a barrier to free development and commerce. But critics observe that socialist societies are not the only ones to have weak intermediate associations, and counter examples like familial China tend to negate this view. Family itself can also limit medium range interactions that enable markets to form from below rather than from above.

Fukuyama's suggestions that corporations are significantly different from governments seems to be another political assumption about democracy, and is doubtful from a promise scaling perspective. This would seem to imply a dissociation of coherent society under imposition (an assumption about the possible modes of government) to be replaced by one based on individual free choice by promises. In practice, few companies are run by free or popular choice; indeed, they may be more impositional than democratic governance models as their mandate to rule does not come from the workforce but is granted by a supreme legal power representing shareholder rights over which no one has control. They are essentially feudal systems.

It might seem that imposition will be a potentially less stable scenario, if followers withdraw their willingness to accept impositions from a leader. In many private organizations, of the kind the Fukuyama suggests are more voluntary, leaders are often selected by imposition and operate by imposition. Workers sign up voluntarily to that kind of model.

1. According to Fukuyama, in Japan, feudal structures lead to relationships between leaders and people (landowners and workers). These could be transmuted into institutions and corporations. In Germany, guild memberships and unions play a similar role.

2. People who do not trust each other may work together under a set of formal rules and regulations as part of an agreement, thus they can trust the general framework. This is consistent with the hypothesis that trust in a coarse entity is easier or more tendacious than trust in individuals or smaller units.

3. Fukuyama does not believe political or corporate activities can create trust. His attitude to religion is unclear, but in Western fashion assumes that religion is based on trust. In practice, church style religion might be more feudal in nature–a prototype of political leadership. The enormous rise of Christianity in Asia is sometimes associated with a more palatable form of society in Western eyes, but this applies to both South Korea and China in equal measure, yet the trust in these societies is still radically different in the eyes of the West.

4. Rules and laws enabled Germany to build post war trust in Fukuyama's view. The ultimate expression of this seems to be, in his view, the unbridled formation of corporations. This has been criticized however. In parts of West Africa there is evidence that corporate activities may have curtailed economic activity rather than be supportive of it.

5. More generally, churches, cults, sports teams, and indeed any special interest membership organization begins to play a role in modern life. People do not share resources with people they do not trust.

Both Putnam and Fukuyama suggest that trust is built up by painstaking local interactions. Government efforts rarely build trust, but may undermine it. Trust is viewed as a cost saver. In Fukuyama's view. widespread distrust imposes a tax on all economic activity.

3.4 Mayer et al, An integrative model of organizational trust 1995

This work is widely referred to as a forerunner in the study of trust [26]. Mayer et al already point out many of the issues about trust and its relationship to trustworthiness and risk, and little progress seems to have been made since. They also point out how trust influences dependencies between agents in a process of extended cooperation. Defining trust as a willingness to take risks places trust and risk on a common scale, aligns with the Promise Theory viewpoint of Burgess [3].

Risk is sometimes also expressed in terms of 'vulnerability', which is a sociological or humanist affectation for risk. Mayer et al point out the confusion with cooperation; there is a semantic difference between cooperation, which might involve trust, and trust itself. The term 'confidence' is also raised as a similar quantity. Confidence is associated with belief, and assessment. Confidence is a prediction or expectation about future behaviour, whereas trust may be based on past behaviour. The link between the two is once again clear but there is a subtle distinction in usage. Trust based on past behaviour certainly informs future expectations of trustworthiness, but not necessarily of kinetic trust.

Mayer et al also discuss the characteristics of a 'trustor' or truster, suggesting that trust is a function of the state of an agent. Is the propensity to trust an invariant characteristic? Trustworthiness is a characteristic of the trustee. These contentions are borne out by more modern neuroscience studies. Trust may be also associated with benevolence (ambiguous) and integrity (promise keeping), which Mayer considers to be universally objective characterizations. This sense of objective morality feels common in the social science literature—based on Christian norms.

Propositions formed by Mayer et al:

1. The higher the trustor's propensity to trust, the higher the trust for a trustee prior to availability of information about the trustee.
2. Trust for a trustee will be a function of the trustee's perceived ability, benevolence, and integrity, and of the trustor's propensity to trust.
3. The effect of integrity on trust will be most salient early in the relationship prior to the development of meaningful benevolence data.
4. The effect of perceived benevolence on trust will increase over time as the relationship between parties develops.
5. Risk taking in a relationship is a function of trust and the perceived risk of the trusting behaviour, e.g. empowerment of a subordinate.
6. Outcomes of trusting behaviours will lead to updating of prior perceptions of the ability, benevolence, and integrity of the trustee.

Proposition 1 suggests an innate quality of agents to assign arbitrary starting levels. In 5, the perceived cost of a risk outcome is an estimate of the damage inflicted by a risky outcome. It's nice to see authors mentioning timescales, i.e. long term effects of trust beyond transactional thinking.

Many qualities are mentioned in this paper but these can be captured under the heading assessments about promises made by agents. Thus Promise Theory remains a sensible candidate for framing these issues theoretically.

3.5 Kumove, Rent-free in your head? how generalized trust is affected by the trust and salience of outgroups, 2023

Kumove provides a useful recent review of the literature, as well as adding some points about groups. Trust between social groups, e.g. ethnic groups, is called outgroup trust. Again this assumption of otherness seems to dominate the sociological literature, but has no obvious root in science. Certain authors argue that, under conditions of ethnic diversity, individuals become distrustful of ethnic outgroups [22]. Kumove notes that, in the USA, outgroup trust and generalized trust do appear to be associated with each other, i.e. that once trust is destabilized by otherness, it also crosses group boundaries. This suggests that trust has a common basis. However, it's possible that US citizens have an unusually small radius of trust. In other studies in Serbian region he finds that "Once again, generalized trust is strongly associated with trust in a range of different outgroups, but the strength of that relationship does not seem to change depending on the salience of the outgroup." He concludes: "Dinesen and Sønderskov (2015) suggested that individuals transfer their outgroup trust into generalized trust by accounting for outgroup salience, but the results here provide no evidence that this occurs." His results pose problems for the existing literature which links ethnic diversity to low generalized trust.

3.6 Glaeser et al, Measuring Trust

In [27], Glaeser et al, discuss the wider concept of social capital and how this could be measured. The authors associate trust with civic engagement and suggest evidence that local government is more successful when it engages with the people from a study by Putnam [23]. By success they refer to economic growth or GDP, judicial efficiency and reduced corruption.

This work reveals the individuality of assessment as much as anything else. Glaeser et al report that different cohorts respond differently to questionnaires. They claim that younger participants are unlikely to say they trust others; older people are more at ease or have simply passed the point of expending energy on the matter. Thus the level of attention (the cost) given to evaluating trustworthiness is a factor in trust offered. The latter seems to be the more relevant point. In other studies the opposite trend has been found.

1. Racial groups (blacks, jews, etc) under stress respond with less trust than elite groups. Furthermore, a higher level of education tends to imply a higher level of trust. This could indicate membership of an elite bubble. The study indicates that marriage seems to make couples more trusting. People between 30-40 tends to be maximally trusting. Only children tend to be less trusting, perhaps suggesting less experience in socialization?

2. "There appears to be some cognitive basis for higher trust. People who say that they have benefitted from the generosity of an anonymous stranger in the past give more. This suggests that trust does not just reflect altruism or risk tolerance, but also beliefs about others which are formed by past experiences." [27]

3. In [27] two experiments are related to trust keeping in terms of promises made and kept. These concern the reciprocation, i.e. repayment of something borrowed.

4. The willingness or unwillingness to make a promise sets a level of trust. The experience over time of the outcomes sampled adjust this up or down.

5. To measure trustworthiness, we use a relationship between amount if money sent and amount returned return ratio—amount returned /amount available to return

   There is great variability in the results, indicating that sociological effects involve great inhomogeneity amongst agents. Trust, in their view, is an incentive to cooperate.

6. In a second game, envelopes containing money are dropped off for participants (like modern click bait) and the addressees are informed. Whether or not they take the bait is considered a measure of trust. Here the measure is only yes or no. The delay in collecting the envelope could be taken as a measure of trust.

7. One should be cautious about attributing trust to single moves in a game. The willingness to engage in tit for tat behaviour or believe in purely altruistic behaviour is known to be an inaccurate predictor of responses in game play. Players might be speculating about future moves in their responses, trying to corral the other user over time. A purely transactional view of the world is clearly incorrect, since trustworthiness is an encoded measure of memory from experience based decisions. Trust is a kinetic form of currency that encodes individual assessments and strategy. It is not a simple function of transactional receipts. This is an important point, which appears to distinguish it from an energy measure. However the latter identification remains to be proven. The obvious counterpoint is game theory in which payoff values are individual, not shared currencies.

8. Racial components and levels of status or organizational membership were identified in some cases: the tendency for someone to send back something of value was less when the participants were of different race. This suggests a tribal or group element to trust.

9. The probability or expectation of meeting again may be different outside a coherence group. (Axelrod)

10. Members of a group tend to inherit a level of trust from their groups memberships. This fits with the hypothesis that the tendency to trust in groups (affinity) is greater than for individuals.

11. This leads to the so-called "free rider" problem, where a single agent can skate along on the reputation of the group.

12. "We find that all of the social capital variables that increase the financial returns for the sender decrease the returns to the recipient. In this experiment these types of social capital lead to redistribution from one player to another. As such, the social capital that we have identified appears to generate private, not group, returns. Findings like this underscore the importance of distinguishing between individual and group-level conceptualizations of social capital." [27]

    The correlation in outcomes of these games was found to be only marginally significant.

3.7 Bergstra and Burgess, Money

In [28], it's pointed out that money (generally economic behaviour) is a proxy for trust. Thus economic ledgers are a simple minded but functional way to assess trust and potentially changing attitudes to trustworthiness. The framing in terms of Promise Theory allows different scenarios to be represented algebraically and graphically in a way not mirrored in other literature.

In certain trust games, the willingness to offer an amount money is seen as a measure of trust, with the amount monotonically related to trust in some sense. It acts as an 'invitation' (as discussed in the business literature), and the trust lies in the assumed promise to return goods worth the amount.

3.8 Mercier and Sperber, The Enigma of Reason, 2017

In their book *The Enigma of Reason, A New Theory of Human Understanding*, Mercier and Sperber argue that human reason is a secondary development after intuitive and emotional assessments of decisions [29]. This suggests that asking people about trust is not the right way to gauge it. People may argue in support of a view which is not directly what they assess by intuition. In this instance, intuition refers to a conclusion reached by purely interior processes.

What is the difference between gut assessment of trust, and a reasoned investigation used for assessment? The latter might be of higher quality, but it goes beyond an kind of initial cache of capital presumed in most studies of trust.

The authors suggest that trust may be considered a self-interest protection strategy to distrust others. Without acceptance of information communication wouldn't be beneficial, indeed pointless. So trust manifests in a practical role as an access control policy. In nature, natural selection is one process by which strategies and decided in the long term, but the process is vulnerable to the flexibility of the algorithm. Adaptations that offer camouflage or advantages to organisms can be imitated by others as a protection, enabling species to lie about their nature. Should predators trust the signals sent by animal markings then? Cheating about semantics is a human concept, but it may be advantageous if one can get away with it even in a non human scenario. There is a framing in Promise Theory which takes these issues into account.

3.9 Robbins, Measuring generalized trust, 2021

Robbins has a recent PhD in social science, and compares two new scales for the measurement of trust with older ones, providing a recent perspective [30]. He writes that:

> "The Stranger Face Trust (SFT) questionnaire and the Imaginary Stranger Trust (IST) questionnaire are two new self-report measures of generalized trust that assess trust in real (SFT) and imaginary (IST) strangers across four trust domains. Both were designed to be objective, empirically valid, and easy to administer and score. Confirmatory factor analysis and structural equation models established the internal consistency, convergent validity,discriminant validity, and criterion validity of SFT and IST. Further tests revealed that SFT and IST correlate with well-established predictors of generalized trust, while other correlates like the age-trust relation were called into question."

Robbins claims that

> Despite the cross-disciplinary breadth of generalized trust, a feature commonly shared by all of this work is a reliance on a small set of survey items(Nannestad 2008). The classic measure of generalized trust is the most-people trust question first developed by Rosenberg (1956) as part of a faith-in-people Guttman scale: "Some people say that most people can be trusted. Others say you can't be too careful in your dealings with people. How do you feel about it?"

The introductory remarks are revealing:

> Despite methodological advances, researchers still use modified versions of questions that were introduced in the 1940s and 1950s... (Bauerand Freitag 2018:20) and, because of that, issues related to operationalization remain. While debates exist regarding the malleability or durability of generalized trust—is generalized trust a function of early life experiences or does generalized trust evolve with lifelong experiences (Bauer 2015; Glanville and Paxton 2007; Paxton and Glanville 2015) most agree that it captures optimism and unconditional faith in strangers and unknown others (Uslaner2002). That is, the notion that person A trusts, period, regardless of the person, the matters at hand, or the conditions in which trust is placed. This conceptualization of generalized trust poses a number of interesting empirical issues related to operationalization.

First, this suggests that social science has a rather specific and moral notion of trust, as noted by Burgess [3]. It's unclear what the methodological advances are that are referred to here. The literature reviewed here basically used variations on the idea of questionnaires to evaluate trust. Some authors commented on the fact that they knew of no behavioural measures that could correlate with trust.

Robbins asks: what is the scope under which trust should hold? How does this vary between cohorts and contexts, places and times?

The trusted face question is interesting for what biologically encoded processes can recognize about interactions between individuals, but also rests on the assumption that the timescale over which trustworthiness varies would be generational rather than something 'in the moment'.

The notion of 'generalized trust' applies to trusting attitudes formed independently of context or trustee. To what extent is an individual oblivious to context in making decisions to cooperate? Is this generalized trust best carried out by person or by domain? In other words are there sufficient commonalities associated with group membership?

Several different contextual semantic scales result from the questionnaire methodology:

- Misanthropy scale (would you trust 'most people', are most people helpful?) with answers yes, no, don't know [31].
- Generalized Social Trust scale (would you trust this person?) - not very much, somewhat, don't know [32, 33]
- Particularized Social Trust Scale (would you trust a person from this group?), with answers not at all, completely, somewhat, don't know [34].
- Political trust scale (how much confidence do you have in this organization?) with answers not very much, quite a lot, don't know [33].
- Betrayal aversion - "If I suffer a wrong, I will take revenge as soon as possible, no matter the cost" and "If someone offends me, I will also offend him or her."

The 'most people' trust question is not sufficient to be predictive of any dynamics. It attempts to take a snapshot of individuals as in the dead brain problem.

Robbins used the term 'discriminant validity', to mean the degree to which measures of different traits were unrelated: social preferences of risk seeking, betrayal aversion, and social desirability were used to assess discriminant validity, based on the other literature.

An additional 'social value orientation' instrument consisted of nine hypothetical decision scenarios, where participants decided for each scenario how to divide resources between themselves and a hypothetical stranger. Each scenario includes three options: a cooperative choice, which maximizes joint gain; an individualist choice, which maximizes personal gain without regard to the other's outcome; and a competitive choice, which maximizes the difference between gains to self and other. Participants were classified as cooperative, individualist, or competitive if they made six or more choices corresponding to one of the social value orientations.

Note the interest in identifying cohorts rather than in the dynamical behaviour "The results indicated that SFT and IST have common predictors: age, gender, religious attendance, party affiliation, and associational memberships consistently predicted SFT and IST. That is, older adults, women, individuals who do not attend church services, Republicans, and individuals without associational memberships have less generalized trust (as indicted by SFT and IST)than younger adults, men, individuals who frequently attend church, Democrats, and individuals who are members of many associations, respectively.Between SFT and IST, the only inconsistent predictors were religious affiliation, religiosity, race-ethnicity, and

region. All other variables, such as income and education, yielded statistically nonsignificant effects for both SFT and IST"

There was only a weak to moderate association between behaviours towards others and moral values. This is hardly a resounding endorsement of the moral philosophical viewpoint as a basis for social These analysis. Worse, relationships to age and other groupings were negative. Some studies have shown that older people tend to be more trusting, others have shown the opposite. It's difficult to relate this to differences in measurement, since the populations answering the questionnaires were also different. The assumption that trust differentiates along cohort lines is thus not consistent with the assumption of the same.

3.10 Rehm and Rahn, Individual-level Evidence for the Causes and Consequences of Social Capital

From the paper abstract [31]:

- Theory: Social capital is the web of cooperative relationships between citizens that facilitates resolution of collection action problems (Coleman 1990; Putnam 1993). Although normally conceived as a property of communities, the reciprocal relationship between community involvement and trust in others is a demonstration of social capital in individual behavior and attitudes.
- Hypotheses: Variation in social capital can be explained by citizens' psychological involvement with their communities, cognitive abilities, economic resources, and general life satisfaction. This variation affects citizens' confidence in national institutions, beyond specific controls for measures of actual performance.
- Methods: We analyze the pooled General Social Surveys from 1972 to 1994 in a latent variables framework incorporating aggregate contextual data.
- Results: Civic engagement and interpersonal trust are in a tight reciprocal relationship, where the connection is stronger from participation to interpersonal trust, rather than the reverse

The paper seems to address questions carefully and critically. The method still uses self-assessment as the criterion for trustworthiness. Confidence in government is also self-assessed ad hoc. In their methodology, interpersonal trust is measured numerically by the usual self-assessment questionnaire: what do you think most people are like? This is consistent with Promise Theory for agents as well as with norms for other authors. The leap directly to psychological factors is telling of a moral subjectivism prejudice.

The authors point out that Putnam's virtuous or vicious cycle of engagement corresponds to the *resonance issue*, as discussed by Burgess in [18]. This is related to Axelrod games also [7, 8, 35, 36].

1. The literature hypothesizes that income inequality would affect trust. "When society's rewards become more inequitably distributed, people may begin to feel exploited by others, thus diminishing their faith in their fellow citizens." This kind of prediction should come out of a model of trust dynamics in a satisfactory model.
2. Fluctuations of valuation and assessment in individuals would be expected to depend on mood (context) and recent happenings. So the way we choose timescales is important.
3. Civic engagement is singled out at a single scale of aggregate engagement. This is indicative of single scale thinking, common in many fields of research. It doesn't stop us from generalizing the idea to other groups. Trust is thus considered to be an interpersonal quality or quantity reinforced by norms. Informed by childhood conditioning in general policy. In other words, trust is a learned quantity.
4. Strong networks enable communities to solve collective action problems by breeding cooperation and easing coordination. Secondary associations such as church groups, labor unions, school groups, and fraternal organizations are especially important manifestations of community interaction. The PTA, for one, invests social capital directly in education.
5. "In and of itself, identification of the aggregate phenomenon at an individual level is an important advance in evidence supporting the social capital idea... As Levi (1996) has argued, there is evidence here that confidence in institutions affects interpersonal trust: our results suggest that social capital may be as much a consequence of confidence in institutions as the reverse."

This is compatible with the coarse grain hypothesis.

3.11 Lewicki and Brinsfield

Suggest that trust is a psychological state. This might be true of assessments made by humans, but it's harder to generalize a patently human notion with the broader cases in technology and other organisms where similar behavioural potentials may play a role.

> Some approaches to trust have described it as only from the perspective of the trustor (Rotter, 1967; Stack, 1978), while others have argued that a full understanding of trust must incorporate the qualities and behaviours of the trustee, or the person being trusted (for example, Mayer et al., 1995). Others have argued that trust is not a single, unidimensional construct; some have argued that trust and distrust are independent constructs (Lewicki et al., 1998) while others have argued that there are different types of trust (Lewicki and Bunker, 1996) and that distinctly different types of trust judgments occur when trust relevant information is processed more rationally' or more intuitively' (Kramer, 1996). Finally, some authors explore how trust changes form and shape as it develops and builds, or as it is broken and declines (cf. Lewicki et al., 2006).

Trust is a willingness to be vulnerable to another party based on both the trustor's propensity to trust others in general, and on the trustor's perception that the particular trustee is trustworthy (Mayer et al., 1995).

Also refers to the Prisoner's Dilemma as a game of trust where payoff is social capital related to trust.

3.12 Bettencourt scaling, coarse grain size and proxy trust

Geoffrey West's approach to scaling in biological organisms are cities, was developed further by Bettencourt. The scaling of communities is a relevant source of data for understanding social measures. The size of a city, where trust givers dwell is known to be related to measures of output. Perhaps some of these could be argued as forms to trust. This is significant because we have models for the scaling of measures in cities [37, 38]. For example, levels of crime rise in cities by size, with superlinear scaling. Crime is at least superficially related to generalized trust.

There is a major difference between the normal concept of trust and the normal concept of energy. Energy is, by its utility, a universal local invariant. It's considered independently immutable, not merely a matter for individual assessment. This is not necessarily an accurate impression of the necessary and sufficient requirements for such an exchange currency.

Why would we assume a conservation law for trust? One answer might be that the postulate simply works, if it indeed does. Another reason is that it might be compatible with a translational symmetry in time. These are equivalent by Noether's theorem, but both questions involve an assumption by formulation. So the question becomes: can we prescribe such a formulation?

4 Trust in Economics

4.1 OECD, Restoring Trust in Financial Markets

This anonymous article [39] seems to be written in the wake of the Enron affair and the first technology bubble. It concerns strengthening trust, but doesn't explicitly address what trust might be.

He notes: Confucius said that a ruler needs three things—weapons, food, and trust; if he has to give up any of these, weapons go first, then food. "Without trust we cannot stand".

Governance is at the root of the argument. He/she suggests that business culture or attitudes may be changing. Information technology means faster access to information that might increase volatility. High speed trading is performed on microsecond timescales. "Meanwhile, the fat and slack which provided some cushioning of volatility in the past has gone. Prudential regulators used to talk approvingly about 'healthy profits' in the banking sector, which the economists took as proof of inefficient 'excess profits/rents'. As these rents disappear, the system becomes more efficient but also more fragile."

> "Now many companies have almost no tangible assets, and their worth lies in the way the various resources (staff, marketing machine, customer lists, brand names) have been brought together. Streams of future earnings are given value, even though the associated work has not yet been done. There is no dispute that these intangibles have value we see this when such companies are sold. But we also see that this value can disappear quickly as markets

> change and evaluations shift. Of course this was always true with assets even physical assets. But the potential for re-writing the accounting script seems much greater today than before, and the accountants have gone from being humble servants of the business to partners in the promotion of the stock."

This latter point is significant. It doesn't tells us what trust is, but it tells us how assessments are the key to defining it. As assessments about which promises are in focus evolve, so does the effect of trust.

What should we do when things go bad? The government should "do something" is a common response. Rules, codes of conduct, and regulations are a common retort. However, there is a tendency to over-regulate at time, strangling business. Parsimony is a theme. We want business to be squeaky clean, but it probably can't be if we want innovation and natural evolution. Evolution is always cut-throat. Some retort that moral compass is missing from business. The implication of morality, as in social science, also seems to be more of a cry of anguish than a plausible explanation, let alone useful advice.

The author argues that self-regulation (voluntary cooperation) is a powerful argument. Reputation, he claims, is a powerful force. No evidence is given for this however; it's taken as a truism.

> "Trust, as Confucius noted above, is a fine quality. But we need to think of this as a balancing-act. Trust obviously enhances and lubricates commerce, so some rules are needed to enhance trust: you need rules and standards (we don't want to have to check the quality of the water supply before we have a drink, or the safety of the aeroplane before we fly). But if trust relies on a third party facilitating or guaranteeing the transaction, then we are into the territory of confused responsibility and moral hazard."

The author argues that protectionism should be limited to cover only absolute necessity, as protections can be abused.

> We used to joke about the idea that depositors, on entering a New Zealand bank, would pause at the counter to peruse the bank's balance sheet before depositing their money, and return at regular intervals (how often? In this fast-moving world, every few minutes!) to check that things were still OK... It is now pretty clear that the "efficient markets" view of the world is a theoretical construct, inadequately capturing reality. If this "efficient markets" view of the world were correct, we would not see: i) Sharp movements in asset prices not associated with any new "news", ii) Regular medium-term swings in asset prices over the course of the cycle, iii) The sort of "irrational exuberance" seen during the Tech Bubble, iv) Failure of uncovered arbitrate conditions to hold in the foreign exchange markets.
>
> Perhaps most damaging for the system as a whole is the market's tendency to foster asset bubbles—bull and bear markets. This behaviour is so far from the text-book model, that some economists simply deny that these price movements are irrational. When prices move down, forces are unleashed that make them move down even more. So markets have great potential for instability, and Milton Friedman's stabilizing speculators are a very rare breed.

The author cites diversity of financial institutions and investors as a safeguard against market fluctuations. He suggests onion levels of risk taking to improve predictability. This segregation by risk is something that bank investment portfolios seem to have adopted in recent years.

It was previously thought that convergence, i.e. making banks monitor everything, was the way forward. Diversity and requisite complexity is easy to argue against.

"If ever there was an old-fashioned idea whose time had come and gone, it is that we need an army of small shareholders to reconcile us with capitalism and keep socialism at bay. Do we really benefit from having everyone watching the stock-market ticker all day and spending their spare time sucking their pencils over their next stock- pick? This is not to suggest that individual direct ownership should be discouraged, but rather to argue that widespread direct ownership does nothing for good governance, other than breeding calls to the political system for more efficiency-sapping rules"

This is basically an opinion piece by an experienced economist of unknown sort. What we learn about trust comes from reading between the lines, within the context of money markets. He ends by quoting Alan Greenspan—once revered as a market magician, and finally declared emperor naked by the collapse in 2008, admitting that he got it wrong.

4.2 Trust and financial markets

This article [40] is hard to judge for its scientific quality, but it is representative of a lot of other work and acts as a summary. It also points out the attention that trust management attracts by major institutions.

It begins with the assumption that public trust is built on the premise that markets serve a purpose that is beneficial to societies, directly in terms of supporting sustainable economic growth, and also indirectly through positive spillovers to other stakeholders.

> "As financial markets are the primary mechanism to intermediate between investors and economic actors, public trust in markets is vital to its role to effectively and efficiently convert savings into productive economic growth, and in turn to reward capital providers with long-term returns commensurate with risks.
>
> the Global Financial Crisis caused public trust in financial markets to decline sharply amid the heavy market losses on both traditional and complex financial products. In response, from policy makers engaged in efforts to craft a coordinated global policy response across affected countries by providing a liquidity backstop for the financial system, highly accommodative monetary policies across OECD countries, recapitalization of core banks and other large financial institutions, and targeted central bank programmes to restore intermediation through markets. Over the post-crisis period, improving market conditions and the continuation of efforts to address the faultiness of the crisis through regulatory reforms have gradually improved public trust. Nevertheless, by some measures it remains fairly low, which calls for policy makers' attention."
>
> A conceptual framework for assessing how trust could impact markets must balance the perspectives of the individual investor and the public at large. In this regard the concept of trust in the markets differs from aspects of investor confidence related to conditions that maximize short-term returns based on assessment of economic and business fundamentals.
>
> For the individual investor, trust may take several forms, including:
>
> * predictability of behaviours (based on historical experience) from markets that are efficient, open, stable and sound, and result in returns commensurate with risks;
>
> * confidence that the rules and oversight of market interactions support the soundness, fairness and integrity of markets;2 and,
>
> * that, both within and beyond the established rules, market participants' behaviours will be ethical in serving the interests of customers.

Associates market integrity with quality of information.

The article discusses crypto assets and their rise and fall, initially claiming transparency and immutability of their ledgers as a security feature. After a number of incidents, trust in these technologies became greatly reduced leading to "stablecoins". These also crashed. "Various analyses of crypto-asset markets highlight challenges including rapid market developments and the fragmented nature of the markets; lack of transparency (including the identity and location of token issuers); and data gaps that hamper proper assessment of risks (FSB, 2018). Moreover, the debate over crypto-assets has drawn attention of policy makers to give further consideration to centralized digital currencies backed by central banks."

"The link between sovereign debt management and public trust is important for the functioning and liquidity of the debt markets, upon which pricing for other traded risk products occur. Principles of sound public debt management are followed to strengthen the international financial architecture, promote policies and practices that contribute to market stability and transparency, and reduce countries' external vulnerabilities."

The article is somewhat vague and superficial, making remarks from a public policy perspective. It seems to measure trust in terms of capitalization of assets, i.e. how much money people are willing to put on the table—corresponding to kinetic trust. That makes financial markets rare. The speculative aspect limits the role of trustworthiness and replaces it by purer gambling of kinetics.

4.3 Huck et al, Pricing and Trust, 2007

The paper uses a game theoretic simulation model to investigate the effect of price regulation on trust. The experiment involves student volunteers.

> "Buyers of an experience good are uncertain about its quality before they buy, but learn (or experience) the good's quality after having bought and consumed it. Experience goods cover the broad middle ground between the extremes of goods involving no quality uncertainty at

all (so-called inspection or search goods) and goods for which quality is not fully revealed even after the consumption (credence goods).

A key role in markets for such goods is assumed by trust. Buyers may buy an experience good if they trust sellers to provide high quality, and will abstain if they do not. In other words, trust induces the demand for experience goods. In contrast, lack of trust impedes mutually advantageous transactions and results in low market efficiency." [41]

With regulated prices firms lose one of their two marketing instruments.

Our finding highlights reasons for why demand may not be downward sloping in markets that suffer from informational deficiencies. In the markets we implement, these deficiencies induce moral hazard but similar mechanisms may operate in case of adverse selection. Regulation may directly aim at removing such deficiencies (for example by introducing standardization, certification, or watch dogs) but in many cases such direct regulation may be very costly. Price regulation is much cheaper to implement, administer and enforce than most other alternatives. Yet, as we see here, it can be very effective.

Remarks "Competition has generally two elements: choice of trading partners and choice of price".

4.4 Interview Alvin J. Huss Professor of Management and Strategy, Kellogg School of Management

Verbatim:

Economists care about trust because it is closely connected to economic activity. Its absence leads to lower wages, profits, and employment, while its presence facilitates trade and encourages activity that adds economic value. For sellers, trust can become a critical competitive advantage: Buyers are more likely to do business with companies that they believe to be "virtuous sellers"—that is, not solely interested in maximizing profit. To compete, profit-maximizing "rational" sellers must use contractual alternatives to trust, which are usually a poor and costly substitute. The result is that virtuous sellers can have an important influence on the market by creating incentives for rational sellers to mimic them [42]. There are really two reasons for why it may be rational for you (the buyer) to trust me (a seller). One is that you may think that I'm what's called a "good-type" or a "virtuous-type seller"—that is, I'm not really a coldhearted homo economicus who, at any moment in time, tries to maximize his profits.

Now, suppose that you know that I am a coldhearted homo economicus; can it still be rational for you to trust me? And the answer is yes, provided that I care not only about today's transaction with you but also about future transactions, either with you or with others.

In deciding whether to honor my promise to you, then, I face a trade-off. By breaking that promise today, I can make more money today, but now it comes at a cost of less future business, essentially.

As long as I care enough about the future, it is then rational for me to keep my promise to you. And since it's rational for me to keep my promise to you, it's rational for you to trust me in the first place.

In summary, then, there are two strands in the literature: one focuses on good types, and one focuses on good incentives. And together, they have a number of implications and shed light on a number of economic issues and phenomena.

5 Trust in Computer Science

In Computer Science and Information Technology (IT), there is a need to automate processes. This includes both processes that lead to trust, with non-human proxies, and methods by which trust can be assessed. Most of the methods used in software are based on the notion of checkpointing and validating of credentials in order to dispense with the process and move into a 'secure zone', analogous of airport security. Once someone leaves a validated zone, their state may be tainted and unfit, but as long as valid credentials have been approved, agents will be assumed safe to interact with.

A technical view of trust, used in Computer Science, associates both kinetic and potential trust with the subject of risk in online interactions. Trust is basically imagined as a token or receipt for an initial

validation of user identity, in the most basic checkpointing which an agent can carry with them during future interactions. Alternatively, in The Clark-Wilson model, trust is attached to the independence of programs that act as gatekeepers to keep privileged promises, e.g. setuid root programs.

In Computer Security, trust products, methods, and other computer professionals, who often refer to Security Theatre in place of real security, is often lacking. This illustrates how trust may involve ego and grandstanding too. The culture in computer science is: can we beat the other guy with a better product or a better piece of software, even a better conference paper.

In Computer Science the hard problems of trust are nearly always deflected onto someone or something else, in order to avoid the problems it incurs. Engineers are often stymied by social concepts and look for rational approximations to paper over the gaps in their models. Rational approaches in turn tend to assume that trusted credentials will suffice to eliminate trust. Credentials only become trusted by trusting in their issuer, however, and thus there's a certain amount of theatrics involved in the claim of trustworthiness by delegating responsibility to 'third party providers' or code bases that are considered to be impartial and infallible.

Computer Science treats obligation and coercion as the standard model ("command and control") of cooperation. There is a pervasive belief that one can dominate and control all systems if sufficient power is applied. Promises and contracts are often presented as weak and untrustworthy attempts to secure cooperation. This points to a weakness in the assumptions of Computer Science, where protected environments are designed into technology for transactional integrity. Engineers will try to make humans behave as machines rather than design machines that can work with humans. It's well understood that the world is not deterministic and that protected environments are easily broken by real world assumptions. Promise Theory seeks to undo that thinking. This leads to many unfortunate surprises as discussed in the context of system safety.

Trust is typically considered to be a weakness in information technology. The term 'trustless' technologies has become popular as a marketing ploy, to suggest that imposing certain rational methods can make systems secure [43, 44]. This means that one trusts one's own assumptions about clients to a service. In many cases, 'trustless' systems rely on opaque systems that may be vulnerable to attack and exploitation, such as blockchain ledgers and third party identity labelling services. The protections they proffer are mostly against outsiders. They do little to secure agents from others within a trust boundary.

In many cases, software developers feel happy to trust these services and the software developers behind them, as they identify with them and think of them as 'their people', in order to feel better about security. Distributed ledgers, like blockchain, may lead to traceability of transactional records, which offers a kind of protection if there are parties with sufficient power to act on misdeeds from the records.

In ad hoc networks, trust is handled by a kind of gossiping protocol that exchanges tokens [45]. Many authors in computer science muddle the concepts of trust and opinion propagation. The tenuous link between these is reputation (e.g. by gossiping protocols); e.g. in [46] the authors proposed a three-condition model for describing the trusting behavior of each user in social networks. There are numerous papers about reputation models, but these are qualitatively different from trust.

Blockchains attempt to induce confidence in technology by exposing data to a diverse group of actors. This seems in line with the entropy trust hypothesis. However, this doesn't mean such a scheme is foolproof of efficient.

This reveals another aspect of trust, relating to power. Powerful agents may be trusted if they are considered aligned with one's own goals, because the power increases their chance of a successful outcome. On the other hand, great power often corrupts agents and promotes self-interest, having the opposite effect on trust. So power touches on a subtlety in trust semantics: reliability versus goal alignment (the latter is often associated with morality)[1].

The fact that one can be accepted simply by showing the right identification is considerably more simplistic than sociological notions of trust, but similar in terms of scientific legitimacy. If one trusts people to give their own assessment of trust, then one might as well trust someone's credentials. This is all 'he said, she said' level assessment.

5.1 Bergstra and Burgess, Local and Global Trust Based on the Concept of Promises

Bergstra and Burgess define trust as the assessment of whether promises have been kept, in the framework of Promise Theory. Promise Theory offers a more formal approach in which to define quantities and

[1] Morality occurs as a possibly misplaced assumption about shared goals and promises. When goals are not shared or aligned, agents may be accused of being immoral by others. See accusation theory by Bergstra [47].

qualities underlying trust, but is currently missing a firm link between sociological and technological attitudes [9]. Network effects of agent interactions can be defined through eigenvector centrality, which accounts consistently for dependencies, but it cannot be computed without long range cooperation, which in turn requires trust to keep promises to share information.

This leads to a bootstrap problem with trust, which is captured by the eigenvalue problem [3]. It suggests that trust is a non-local phenomenon that can only be post hoc consistent. We therefore have to take seriously the idea that local fluctuations will dominate dynamics over short timescales.

5.2 Jøsang, Trust and reputation systems

Audun Jøsang is one of the active figures in IT security research, who has written papers alongside and contemporary with [9] over the years. Here he provides a review and critique of IT security notions of trust. It's hard to trust citations and references in IT research, as authors do not typically cite one another unless there is some benefit to them, so a review by someone who works on security full time is a useful data point.

Jøsang points out that fraudulent behaviour on the Internet became a problem principally with the opening up of the Internet to commerce. Before that, the smaller select group of academics were more predictable and formed a tribe of like-minded individuals. Thus the sociological notion of "other" was first introduced by commercial developments—a form of Internet globalization.

> "It is normally assumed that information security technologies, when properly designed, can provide protection against viruses, worms, Trojans, spam email and any other threats that users can be exposed to through the Internet. Unfortunately, traditional IT security technology can only provide protection against some, but not all online security threats." [48]

Unlike nation states, online interactions are not governed by formally ratified laws: "What constitutes ethical norms within a community will in general not be precisely defined. Instead it will be dynamically defined by certain key players in conjunction with the average user." Jøsang further notes that: "The traditional cues of trust and reputation that we are used to observe and depend on in the physical world are missing in online environments. Electronic substitutes are therefore needed when designing online trust and reputation systems."

The paper concerns itself mainly with trust management, overlapping sociology, psychology, computer security. Matt Blaze was the one who pioneered trust management in the 1990s. The paper defines trust mainly under the heading of reliability, which is consistent with [9]. "Trust is the subjective probability by which an individual, A, expects that another individual, B, performs a given action on which its welfare depends." This is consistent with the expectation of promise keeping referred to in [9]. Trust is later defined as a willingness to *depend* on another individual or entity to achieve an outcome. This can also be called a risk appetite. This introduces a networking aspect into trust, which fits with the network centrality view of [9].

Jøsang seeks to fold IT security into the general framework of trust, risk, and safety, but doesn't attempt to review the sociological literature on these at all (again this is normal practice in computer science). The definitions are somewhat handwaving and ad hoc, which is also normal for non-formal methods. Like many other authors, he argues for a measure based on probabilities. Well founded trust is one for which one knows the probability. Probability is a seductive method for scientists and engineers, because it seems to solve the problem of quantitative values. In fact, it can be misleading as it covers all its tracks and lacks the dimensionality to be a faithful representation of the original data.

Jøsang also argues that trust should be *transitive*, in line with many other authors, i.e. if A trusts B and B trusts C, then A would trust C. While this is a desirable idea from a modelling perspective (it simplifies logical inference and algebraic rules, for instance), it feels more like wishful thinking. Promise Theory predicts that this idea as simply incorrect, as noted by Bergstra and Burgess [9, 10].

More interesting, are the Hierarchical Reputation Systems discussed: Slashdot uses moderator hierarchies, as does Wikipedia. These are trusted third parties within a closed system of assessment. the hierarchy consists of service users, moderators, controllers. This is like the justice system, with judges at the top of the pyramid.

Industry standards and peer trust schemes have also become an industry within IT. The so-called Trusted Third Parties (Verisign etc), proclaimed their objectivity. Then companies looking to dominate the space (Microsoft and IBM initially, then Google and Facebook) created standards which they hoped would bring customers to them. Jøsang notes that functionality and flexibility may suffer by rigid mechanisms.

> In identity management, the term Circle of Trust is defined by the Liberty Alliance to denote a group of organizations that have entered into an agreement of mutual acceptance of security and authentication assertions for authentication and access control of users. The Liberty alliance has adopted SAML2.0 [42] as the standard for specifying such security assertions. The WS-Trust standard 5 which has been developed mainly by IBM and Microsoft specifies how to define security assertions that can be exchanged with the WS-Security protocol. WS-Trust and WS-Security have the same purpose as, but are incompatible with SAML2.0. It remains to be seen which of these standards will survive in the long run. Other trust related IT terms are for example
>
> - TTP (Trusted Third Party), which normally denotes an entity that can keep secrets
> - Trusted Code, which means a program that runs with system or root privileges
> - Trust Provider, which can mean a CA (Certificate Authority) in a PKI.

As with sociological measures, IT usually treats trust as being a discrete (binary) issue. Either you trust or you don't. Users are even pressured into signing off on trust completely, without recourse or doubt, for example when accepting the identity of a foreign credential. This becomes a game of ticking boxes rather than taking risk seriously.

Jøsang discusses the use of 'second opinions', which is an interesting twist on reputations. Risk trust matrices based on fuzzy logic and ad hoc functions to represent probabilistic thresholds [48–50]. The temptation to use logic and probability as crutches for reasoning is strong in IT. In other work found on Wikipedia, Jøsang adopts a so-called Opinion Model, based on subjective logic, may be a suitable technique for assigning trust values in the face of uncertainty. An opinion is a representation of a belief and is modelled as a triplet, consisting of: b (a measure of one's belief), d (a measure of one's disbelief) and i (a measure of ignorance); such that $b + d + i = 1$. It is assumed that b, d and i are continuous and between 0 and 1 (inclusive), so this has the appearance of convex game-based modelling. The weaknesses are already known, yet researchers tend to support such approaches in spite of their weaknesses, since they offer a form of 'science theatre':

> This model's strength reputedly lies in the ability to reason about the opinions (on a 'mathematically sound basis') and its consensus, recommendation and ordering operators. However, its major weakness is that it cannot be guaranteed that users will accurately assign values appropriately.

Jøsang concludes that trust and reputation systems are vulnerable to attack. This is an interesting claim, which is surely mirrored in any measure of societal trust. It's a particular affectation in IT that assumes doubt can be replaced by Boolean certainty.

> Any reputation system with user participation will depend on how people respond to it, and must therefore be designed with that in mind. Another explanation is that, from a business perspective, having a reputation system that is not robust can be desirable if it generally gives a positive bias. After all, commercial web stores are in the business of selling, and positively biased ratings are more likely to promote sales than negative ratings.

5.3 Graydon, An Investigation of Proposed Techniques for Quantifying Confidence in Assurance Arguments

Graydon contends that probabilistic approaches to trust are as flawed as any other approach. Trust is related to confidence, but probability doesn't necessarily measure that more reliably than ad hoc assessments. The semantics are similar. In [49], the authors intend to help analysts to make accurate assessments of confidence in other parties, using a form of Bayesian Belief Networks. This is similar to the idea of machine learning. They are sceptical of using statistical methods with quasi-quantitative measures.

> A proponent might propose that the need to assess confidence justifies using one of the available quantified confidence techniques even though its efficacy is not firmly established. But a quantitative technique that is no more effective than qualitative techniques might be riskier: some readers might come away with a greater impression of the trustworthiness of the quantitative analysis than they would a qualitative analysis is about which equally little was known. To use quantities of unknown quality is to risk committing the fallacy of over

> precision. That is, the appearance of precision given by quantities might lead readers to put undue trust in confidence assessments

Probability is a standard tool in science, but its validity and its semantics are not always easy to impute:

> We do not claim that our results show that the probability theories underlying the proposed techniques are in error. However, the proposed uses of those theories—the techniques we review—are, at best, imperfect. Without research that provides strong, direct evidence that the resulting confidence assessments are trustworthy, there is no plausible justification for relying on one of these techniques in making decisions about which critical systems to deploy or continue to operate.

It's rare to find papers critical of methodology in this way. Usually people are clutching at straws to make any kind of inference, so it's refreshing to see an honest attempt to appraise the techniques.

5.4 Noorian, The State of the Art in Trust and Reputation Systems: A Framework for Comparison

This review discusses the work of a number of other authors [50]. It touches on similar issues to the previous review: individual trust, subjectivity, group reputation, social dimension to capture trustworthiness, etc.

Yu and Singh deal with deceptive agents who deliberately disseminate misinformation through network for their self-interest. The proposed model considers three types of deceptions: complementary, exaggerative positive and exaggerative negative. The classification is based on the behavioral model of the participants in giving ratings. Remarks:

- They hint at Role Based Access Control (RBAC) provides a mechanism for context. This is a common industry standard, so it's good to see it mentioned.
- These authors also assume transitivity of reputation.

5.5 Wong et al, Machine Learning and Trust

This is part of a series of papers about using machine learning to identify behavioural data patterns that signify and quantify trust has begun recently [51, 52].

This work is perhaps closest to the generalized anomaly detection approach in this NLnet trust project. The chief goal is to gauge the trustworthiness of deep neural networks for pattern recognition. It has implications for technology in the future as well as attitudes to trust. Curiously, the idea of training a machine to determine trustworthiness by copying humans so that humans don't need to do this themselves is slightly paradoxical.

> Question-Answer Trust: The function used to measure a model's trustworthiness for a single question-answer scenario in a human interpretable manner. The function is defined by $Q_z(x, y)$ which takes a question x, a model M's answer y to the question x, and an oracle O's answer z to question x.

In other words, the trust model concerns the training of a function based on contextualized data. Training is performed by comparing with data from an oracle, i.e. a human source of truth. Thus the training assumes that the human oracle is completely trustworthy. This is an interesting but also paradoxical idea. Trust is that measure of acceptance in something without verification, but here one goes to enormous lengths to train a verification method to answer a question that the human oracle can already answer–just not at scale. This is typical of IT concepts of trust. It illustrates how IT needs to solve the problem of scaling interactions through automation.

The trust model is based on how an agent answers certain questions. The authors seem to have in mend the idea of a question as an image recognition categorization: is this a cat, or a car, or a telephone, etc. They define a trust matrix with all the recognizable objects as rows and columns, then look at the reliability of recognition. Does the recognition algorithm recognize the object faithfully? A fully diagonal matrix would be completely reliable. Off diagonal elements correspond to misidentifications. The non-diagonality of the matrix measures the scatter or lack of trustworthiness or overconfidence of the algorithm.

From a review [53]:

> Generative Pre-training Transformer (GPT) models were first launched in 2018 by OpenAI as GPT-1. The models continued to evolve over 2019 with GPT-2, 2020 with GPT-3, and most recently in 2022 with InstructGPT and ChatGPT.
>
> All GPT models leverage the transformer architecture, which means they have an encoder to process the input sequence and a decoder to generate the output sequence. Both the encoder and decoder have a multi-head self-attention mechanism that allows the model to differentially weight parts of the sequence to infer meaning and context. In addition, the encoder leverages masked-language-modeling to understand the relationship between words and produce more comprehensible responses.

5.6 Verescha, How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies, 2021

The authors discuss how the business world is beginning to look to AI to assess trustworthiness in relationships. The literature does not yet provide guidelines that support the empirical study of human trust in AI-based decision support systems [54]. The main focus is, thus, investigating how to study trust rather than factors that affect trust. In this paper, it's not always clear whether the authors are talking about trust in AI systems or trust between humans determined by AI systems.

> Three main findings are 1) the three theoretical elements of trust, vulnerability, positive expectations, and attitude are not fully integrated in the reviewed papers' experimental protocols and qualitative measurements. There is therefore a risk that some empirical studies capture constructs other than trust; 2) a large variability among the designs and measurements used to assess trust which can impair validity and replicability; and 3) the challenge of investigating the dynamics of trust considering the constraints of laboratory experiments and the applicability of existing methods.
>
> In the rest of our paper, we rely on the Lee and See definition when referring to trust: "An attitude that an agent will achieve an individual's goal in a situation characterized by uncertainty and vulnerability."

The authors conclude that one cannot have trust without positive expectations. Without positive expectations, it is more appropriate to discuss distrust. The latter point contradicts the Promise Theory in [3], where trustworthiness is about consistency regardless of whether an agent is consistently good or bad. Knowing what to expect (predictability) is the key to trustworthy assessment in the Promise Theory view. This is different from accepting someone's promise to perform a critical action. Kinetic trust therefore attaches to specific promises, while potential trustworthiness might be informed by multiple histories.

They make an interesting observation about assessment, however. A position of vulnerability by the recipient (-) might increase the hope or level of affinity to trust in a promise (+). e.g. we really need to sell the house and we trust you to offer a good price, or I need a cure so I trust that your medicine will help me. Agents might feel differently if they were less desperate.

In a rare standout, the authors discuss group effects on trust, and whether groups are more or less trusting than individuals. The claim "repairing trust has been found to be more difficult for groups than for individuals" [20].

> "In recognition of the simultaneous importance and fragility of trust, a burgeoning literature on trust repair has emerged which investigates repair from the perspective of accused parties (trustees) and from the perspective of their perceivers (trustors). When trust is violated, the positive expectations regarding the trustee(s) decrease or become negative."
>
> "in organizations, trust violation and repair episodes frequently involve a social element, including cases in which groups evaluate individuals, which is the focus of our study. Research on organizational gossip reveals how both trust and distrust can be spread and amplified, for example, as subordinates discuss among themselves how they should react to a manager's violation of company policy or industry standard, teams discuss how they should react to a member's alleged free-riding, and supervisors meet to discuss how they should react to a subordinate's alleged incompetence"
>
> "for matters of integrity, trust repair is encumbered when trustees offer an apology rather than a denial, given that an apology would then be considered to offer a reliable signal that one lacks integrity (i.e., highly diagnostic negative information about integrity) which

should outweigh any positive effects from this response's signal of intended redemption (i.e., less diagnostic positive information about integrity). Thus, this research suggests that it is relatively "effective" to repair a competence-based violation with an apology and an integrity-based violation with a denial, whereas it is relatively "ineffective" to repair an integrity-based violation with an apology and a competence-based violation with a denial. Indeed, the robustness of this finding (across settings, despite evidence regarding innocence vs. guilt, and for different types of populations) (Ferrin et al., 2007, Kim et al., 2004) suggests that this interaction may also be found when trustees attempt to repair trust with a group. That is, we expect that a trustee's attempts to repair trust with a group will be less effective when they deny culpability (rather than apologize) for competence-based violations or apologize (rather than deny culpability) for integrity-based violations, just as these responses have been found to be less effective when directed at solitary individuals."

Hypothesis 1: Groups will exhibit less trust than individuals after an ineffective response, but greater trust than individuals after an effective response, has been offered for an alleged transgression.

Hypothesis 2a The lower (higher) trust of groups relative to individuals will be mediated by the expression of persuasive arguments.

Hypothesis 2b The lower (higher) trust of groups relative to individuals will be mediated by the extent to which members feel normative pressure.

Hypothesis 3: there is feedback between group and individual levels.

The paper's hypotheses were tested through interviews. In organizations, judgements are often made by groups. There are resonance phenomena within the group, not just between leader and follower, but forming an echo chamber around "us versus them" issues that are not intra-group but inter-group dynamics. "Both Hypotheses 3a and 3b predicted that individuals would revise their trust assessments when responding later as part of a group (the individual group condition). This tendency is supported by the significant repeated measures MANOVA results for the "individual first" condition. The comparison of groups (at time 1) and groups (at time 2), then, further corroborated this finding."

Allegations and accusations play a role too.

Pointing to reference [55], which states:

> Much research suggests that having a group facilitator can improve group decision-making effectiveness, because a facilitator can enforce a more structured process. Thus, a number of GDSSs were designed to support or even replace some of a human facilitator's roles. For example, in distributed groups, a lack of leadership is often found to inhibit the groups' capability to organize and reach consensus.

They discuss the role of 'embodiment' in forging trust. This could be interpreted as a single identity for a group—a face for the decision process, by avatar in this case. A group leader may help people to focus, but there is no obvious connection to trust.

A group effect is the sense of "always being there", i.e. a continuity through redundancy of a group.

6 Trust in neuroscience literature

More recently, trust is an issue that has been discussed in neuroscience, with more obvious physical measurables. The main things we learn from this are:

- Trust is individual to each agent (person).
- Trust is the result of a process of assessment, based initially on irrational or emotional ("system 1") responses.
- Rational arguments may be used to adjust attitudes to trust ("system 2").

Neuroscience is studied using electro-chemical as well as genetic and brain scanning technologies. These can be correlated with sociological assessments and with game theoretic models.

The so-called Trust Game, a neural variation on the Prisoner's Dilemma model, has been increasingly used in the emerging field of neuro-economics to study brain mechanisms underlying decision-making in social interactions with financial outcomes.

In an opinion-review piece in Trends in Neurosciences, Cell Press Reviews [56], Kreuger and Meyer-Lindenberg looks at the NPE model, or Neuro Psycho Economic view of trust in which trust builds on genetic predispositions (long term memory) and hormonal stimulation (oxytocin and testosterone act as the antagonists of trust) and is enhanced by learning in social networks. The authors also use the classical Prisoner's Dilemma model to discuss the gaming or reinforcement learning of trust, to encode the economic arguments.

> "At the behavioral level, the two-person reciprocal trust game enables to measure both the propensity and dynamics of interpersonal trust behavior. At the psychological level, psychometric and survey measures allow evaluating the psychological systems (i.e., motivation, affect, and cognition) and their linked T-R-U-S-T components (Treachery, Reward, Uncertainty, Strategy, and Trustworthiness). At the neuro-functional level, complementary neuroimaging methods (e.g., functional magnetic resonance imaging, electroencephalography, and focal brain lesions) identify the domain- general large-scale brain networks (i.e., activation and connectivity patterns) shaping the psycho-economic components of trust behavior. At the neuro-chemical level, pharmacological manipulations of neuro-peptide hormones [e.g., oxytocin (OT)] and steroid hormones (e.g., testosterone) as well as neurotransmitters (e.g., dopamine) reveal the neural signaling pathway mechanisms invoked in trust behavior. At the neuro-genetic level, twin and genetic studies looking at individual variations in the human genome and variants of single-nucleotide polymorphisms (e.g., OT receptor gene) explain mechanisms of heritability and genetic variation in producing individual differences in trust behavior."

Zak also identifies a correlation between oxytocin (pleasure) and trust in organizations, thus suggesting the trust is associated with positive feelings [57]. He implies that the causal factor is trust, but can't show this. The emotional factor in reasoning and trust suggests that positivity may also lead to trust or that the two are mutually reinforcing. In a similar vein to Mercier and Sperber, they imagine that trust begins with autonomic responses and is then justified through reasoning using a bounded rationality type argument.

> "The trustor is economically motivated to adopt a strategy to reap context-based benefits, thereby removing uncertainty by transforming economic risk of treachery (i.e., losing monetary stakes) to the extrinsically positive expectation of reciprocity. If trust is motivated by intrinsic incentives (i.e., other-regarding interest), it becomes a socially rational choice, contributing to the relationship success and valuing group belonging. The trustor is socially motivated to evaluate trustworthiness to promote relationship-based benefits, thereby removing uncertainty by transforming social risk of treachery (i.e., being betrayed by the partner) to intrinsically positive expectations of reciprocity."

And reinforcement learning adapts responses.

> "Based on a systems neuroscience view [58], trust arises from the interactions of psychological systems (i.e., motivation, affect, and cognition) that engage key regions anchored in domain-general large-scale brain networks: reward network (RWN), salience network (SAN), central-executive network (CEN), and default-mode network (DMN). The motivational system of trust involves the RWN to determine the anticipated reward for trusting another person. The affective system of trust engages the SAN to incorporate aversive feelings associated with risk of treachery by another person. The cognitive system of trust involves, on the one hand, the CEN (i.e., cognitive control system) to adopt context-based strategies, and on the other hand, the DMN (i.e., social cognition system) to evaluate relationship-based trustworthiness for trusting a partner"

The role of the amygdala is commented, as the emotional detector of treachery.

The social cognition system for assessing trustworthiness is anchored in the Default mode network (DMN), including crucial brain regions such as the temporo-parietal junction (TPJ) and dorsomedial PFC (dmPFC) [13]. The DMN has been consistently identified in the context of mentalizing about others to facilitate cooperative decision making

> "The TPJ (temporoparietal junction) has been linked to various social cognitive functions, including self-other distinction, perspective taking, and intentional inferences of others [40], making it an essential region for inferring and attributing the intentions of others to evaluate relationship-based trustworthiness. Trustors with higher perspective-taking tendencies

show not only greater trust toward others but also reduce their trust more drastically after treachery by their counterparts. TPJ activity increases with age when continuously trusting another person, indicating a higher sensitivity and orientation toward other people's social signals. Sophisticated trustors show higher TPJ activity than naive ones, consistent with the assumption that sophisticated trustors build better mental models about the intentions of their partners models that build not only on what trustees reciprocate but also on what they expect from the trustors regarding initial investments."

"The NPE model describes how interpersonal trust evolves through repeated interactions: from calculus-based trust, through knowledge-based trust, to identification-based trust [44]. First, the trust relationship begins with calculus-based trust. Driven primarily by SAN (risk of treachery), trustors encounter ambiguous situations and perform rational calculations of the costs and benefits of creating and sustaining a trust relationship."

They remark that these findings focus on a particular model of trust in an economic reward setting, and might not cover all cases. The authors summarize:

Behavioral level. Can trust be measured with greater ecological validity by combining qualitative methods (e. g., case studies, ethnography studies, and interviews) with quantitative methods (e.g., exchange games, scale-based trust surveys, and implicit association of trust assays) to link trust measures from field observations, laboratory experiments, and real-world social interactions?

Psychological level. Apart from 'horizontal' trust among individuals, how do the proposed trust components impact the building, maintenance, and repairing of 'vertical' trust between individuals and those holding institutional positions?

Neurofunctional level. Shifting from standardized univariate analysis techniques (i.e., location-based approach) to more sophisticated multivariate ones (i.e., function-based approach), how is the functional (temporal) and effective (directional) connectivity within and between key regions of domain-general large-scale networks dynamically shaping trust over time organized?

Neurochemical level. What are the causal relationships among exogenous administration and endogenous levels of neuropeptides (e.g., OT, arginine vasopressin), sex hormones (e.g., testosterone, estrogen), and neurotransmitters (e.g., dopamine, serotonin) in modulating interpersonal trust? Neurogenetic level. In contrast to gene-specific candidate-driven studies focusing on multiple variations of single-nucleotide polymorphisms (e. g., OT receptor gene), can genome-wide association studies identify genome-wide sets of genetic variants associated with interpersonal trust? Nature/Nurture/Culture. Combining quantitative/molecular genetic studies with trust measures in different cultures, what are the individual and conjointly genetic, environmental, and cultural influences on trust behavior?

Health/Disorder. Applying a computational psychiatry approach, can computational models be built to bridge the explanatory gap between the proposed neuropsychoeconomic basis of trust and the neuropathology that underlies trust impairments in psychiatric diseases?

7 Summary of trust in management literature

Management studies are an area where little science is involved, but a lot of experience goes into subjective formulations. We can take these as indicative of general attitudes towards trust that have been formulated over many years. They are not measurements in a scientific sense, but nonetheless valid anecdotal guiderails.

The Roffey Park Institute suggest that the 8 behaviours that build trust

1. Sticking to commitments.
2. Demonstrating trust.
3. Being personal.
4. Being consistent.
5. Appreciating others.

6. Listening well.
7. Demonstrating vulnerability.

The Franklin Covey organization suggests thirteen behaviours of high trust leaders

1. Talk Straight.
2. Demonstrate Respect.
3. Create Transparency.
4. Right Wrongs.
5. Show Loyalty
6. Deliver Results
7. Get Better.
8. Confront Reality
9. Clarify Expectations.
10. Practice Accountability
11. Listen First.
12. Keep Commitments.
13. Extend Trust.

These are, in principle, assessable. Where does trust originate? From within, or by exchange transaction?

The diffusion of trust through organizations has been argued. but not modelled in [59]:

> "The increased sensitivity to trust issues has caused many companies to embrace the concept of transparency believing it will lead to increased levels of trust among the general public, more specifically investors, mar- ket intermediaries, consumers, and regulators. It is frequently intimated that transparency leads to a more effective organization; and as a result, many companies have redesigned their corporate communications and public affairs departments to enhance their responsiveness to stakeholder demands."

Opinion dynamics about how opinions spread uses diffusion models on random network, comparing to small worlds and scale free networks [60].

7.1 Mezick and Sheffield, Inviting Leadership

This book effectively picks out the idea of invitation as a trust building protocol, using Promise Theory arguments amongst other experiences [61]. It's interesting because it tied together the relationship between trustor and trustee as a resonant interaction, as described in [3, 18], before the analysis was done. In particular the difference between invitation (which establishes a process that tends to increase trust) versus imposition (which establishes a precedent that tends to undermine trust).

7.2 Kleijn, Trust in Governance Networks: Its Impacts on Outcomes

Kleijn discusses an interesting web based study [62], looking for some specific relationships. "We are firstly interested whether trust, as an independent variable, influences [certain] outcomes. We are also interested in the factors that influence trust. Two main factors are our focus in this article: the complexity of the issue dealt with in governance networks, and the managerial strategies employed. We assume that: a) when the issue is more complex, trust is more important, and b) when network management is more intensive and more strategies are employed in a governance network,the level of trust is higher."

The authors identify some key dimensions to trust:

1. Agreement trust: The parties in this project generally live up to the agreements made with each other

2. Benefit of the doubt: The parties in this project give one another the benefit of the doubt
3. Reliability: The parties in this project keep in mind the intentions of the other parties
4. Absence of opportunistic behavior: Parties do not use the contributions of other actors for their own advantage
5. Goodwill trust: Parties in this project can assume that the intentions of the other parties are good in principle

"The Cronbach's alpha of these five items is 0.73, indicating that they can be seen to form a single 'Trust' scale. The items were recoded, added up and divided by 5. Thus, a higher score on this scale implies a higher degree of trust. The mean score on the scale is 3.47(standard deviation 0.56), implying a moderate degree of trust between the partners."

7.3 Venture capital investment scenarios

There is an obvious power imbalance between most entrepreneurs and investors. Investors seek control by trying to add clauses in their favour in contracts [19]. The study shows that entrepreneurs have to trust quite a lot, but less trusting entrepreneurs with leverage may omit clauses that favour investor asymmetry of control. Thus trust amounts to acceptance of external control and additional cost to self. The study seems flawed in that it can't distinguish the power imbalance between entrepreneurs, and it's unclear how the promise of money weighs up against the promise to accept additional controls. Trust seems to be a wildcard here, rather than a deterministic factor.

7.4 Supply chains and trust

Several references study the problem of trust in supply chains. Many of these are studies by Chinese universities driving a prototypical literature on network analysis [63]. A more domain centric but popular review contains some useful insights [64]:

> "a series of case studies looked at the what, why, and how of trust. More than 50 in-depth interviews were conducted across the supply chain: 14 retailers, 13 finished goods assemblers, 12 first-tier suppliers, three lower-tier suppliers, and nine service providers. (Participants were picked mostly on the strength of their reputations as leading-edge supply chain implementers.) A structured interview protocol was used to assure that the answers could be compared while allowing the flexibility to probe unique practices. When managers identified trust, or the lack of trust, as a bridge or barrier to supply chain success, the researchers asked them to define trust and describe the nature of their supply chain relationships."

> "Roughly one in four of the interview participants took pains to point out that real trust is rare in supplier-customer relationships. Others placed trust on their wish lists. Many managers said that the word 'trust' is overused, misused, and frequently abused. The survey results supported those sentiments. Fewer than half of all respondents (49.6 percent; mean=4.37 where 7=strongly agree) felt that trust characterized supply alliances. Worse: Only about one in three managers reports that value-added resources are shared among supply chain partners (36 percent; 3.95). Similar percentages said that their alliances operate under principles of shared rewards and risks (32 percent; 3.74) and that their companies dedicate resources to help suppliers improve their own capabilities (29.6 percent; 3.67). The manifest lack of trust explains why companies place so much emphasis on complex contracts, detailed confidentiality agreements, and specific continuous improvement clauses..

> "In a world where realism may not always be valued but where failed promises destroy trust, companies need to adopt the mantra, "Do what you say you are going to do the first time, every time, all the time!" As one manager explained of a favored supply chain partner, "I never lose sleep when I work with Jim because I know that he will deliver as promised even if he has to lose sleep." Performance is the bedrock on which lasting trust is built...

> "Perhaps the best indicator of trust-based behavior is the true sharing of risks and rewards. Several managers at buying organizations explained that they work diligently to treat suppliers well and that the benefits of joint initiatives are shared 50-50 for the first year."

The stories reveal how supplier relationships are dropped in a heartbeat if buyers can win a few price points advantage. Whether it's a buyer or a seller's market is relevant to who can attempt to exploit the other's weakness. This is a matter of strategy: trust does not mean that the counterpart will be kind, only predictable. If they predictably try to exploit the other, then they are still trustworthy in that context. This is the flaw in thinking that trust is a question of moral behaviour.

> "Trust is defined as an expectation that the other party can meet the transaction requirements in the future cooperation between the subject enterprise and the object enterprise in supply chain cooperation. In this paper, the trust between supply chain nodes is integrated into the network model with edge weights. In the process of constructing the evaluation index model, we refer to the relevant literature at home and abroad, and strive to seek a reliable basis for the trust evaluation of supply chain partners. By analyzing and summarizing relevant literature and questionnaire survey, it is concluded that there are 4 primary evaluation indexes and 16 secondary evaluation indexes." [63]

The difference between rumour, reputation, and individual trust comes to light in these remarks. The review suggests that, in the cut-throat business environment, immediate pragmatism and cost cutting override trust. "Reputation is inversely proportional to the probability of taking opportunistic actions." The importance of products versus the trust in the supplier.

This follows the basic principles of the promise model, with node degree, weights assigned by trust and efficiency of keeping promises. It's not clear that the expressions used actually model what is claimed, but the definitions are what concerns us here. We would expect this to match the centrality arguments in [9]. The role of the downstream principle [10] in supply chains remains to be assessed. There doesn't seem to be any literature on this topic.

8 Examples of trust, anecdotal cases

These notes are reminders, not meant to be complete or encyclopaedic.

Example 1 (Why trust this work?) *In order to be trusted as an author of something from which a community has to lose, we need to offer some compensation in return. If an agent is perceived as seeking to undermine another's livelihood, it will automatically generate mistrust. The invitation move of offering something valuable unconditionally up front will be the spoonful of sugar to help the medicine go down.*

Example 2 (Trust in the trust literature) *The trust literature is broad and vague. Promise Theory suggests that it should be easy to trust it. On the other hand, when looking more closely, many references promise more than they deliver, so when one sees claims of studies explaining trust one assesses that the claims are less than than perfectly trustworthy.*

Example 3 (Rules and boundaries versus accountability) *Another aspect of prioritization is to limit the scope of what needs to be assessed, by constraining the interacting system with rules and boundaries. This includes checkpoints for bounding and modularizing assessments.*

Fast food is possible because we constrain the menu to four meals, without dialogue. It makes the process transactional. Without these rules, agents may have to engage in lengthy back and forth to get what they want. Society is about constraining one freedom to pay for another. This is how to rule society (e.g. pax Romana, violence at the edge to keep peace at the centre).

Trust in the framework allows us to trade currencies of different kinds: risk for time spent.

Example 4 (Bureaucracy, milestones, and accountability—checkpointing) *The prevalence of bureaucratic methods of government and corporate dealings are a hedge against trustworthiness. e.g. EU grants are only awarded on the condition of imposing large bureaucratic burdens of documentation. Imposition reduces the trust in EU, according to Promise Trust.*

Jumping through hoops is an example of 'checkpointing', as one seems at an airport, and is symbolic of an absence of trust (often argued as requiring accountability). Trust is a resource saving strategy. It has the status of a policy, i.e. a formalized intention. When trust is absent, procedures like bureaucratic process may be used to tick boxes and generate step by step confirmation, which takes time and effort. Bureaucracy is a symbol of absent trust.

The risk of non-delivery and threat of legal accountability is considered more important than wasting the finite resources of the agent. This comes about by working on behalf of society, by proxy. The proxy can't offer any promises on behalf the applicant, so it can only promise to verify and punish failure.

Trust is precisely important as a cost saving method because agents have finite resources and need to prioritize, like the tragedy of commons. In Venture Capitalism, VCs gamble on the authority of their investors as a service. There's no reason why model couldn't be used for government too, but it's perhaps harder to sell.

Example 5 (Logos and brands—hacking trust) *Logos are simplified cognitive symbols. They take less effort to recognize than long descriptions. Based on the hypothesis that trust is a cognitive cost saving strategy, they are strategies to hack trust. Trust in cost saving Symbols save on cognitive effort*

We seek to simplify identity. Simplified Chinese versus Traditional Chinese. Proxy codes for product descriptions. These are examples of data compression.

See explainability below.

Example 6 (Repetition, marketing, propaganda, rote) *As a learning process, repetition is trust inducing because it trains and agents memory to be familiar.*

Example 7 (Trust people to follow rules—propaganda) *Can we trust people to follow rules? Rules are generally imposed rather than invited through careful protocols. As* impositions*, we would expect them to fail much of the time, unless overwhelming force could be wielded to coerce. So this probably depends on whether people trust the rules to improve lives.*

If the rule is vague or simple, PT suggests it might be more easily trusted than if it is very precise and conditionally specific, making it less applicable and therefore less frequently relevant.

PT predicts that repetition breeds familiarity and is thus a way to hack trust. Clarity and consistency or reliability in communicating makes repetition simpler and more reliable. If the rule is complex or too specific, PT suggests that it may not win as much trust. In general it takes a lot of campaigning, advertising or propaganda to bring complex or unfamiliar rules to people's attention.

A single source of consistent messaging is key to calibrating the intent of rules. PT predicts that the passing on of a message by gossiping and rumour will tend to align the message with the receiver and sender during the process, leading to randomization.

For example, during the 1970s the UK was exposed to government campaigns "Green Cross man for safety in crossing the road", "Don't drop litter campaign", "Wear seatbelts campaign", etc. These laws eventually became norms, but some now seem to be fading from consciousness since we take them for granted. People may need to be reminded continuously of their obligations. This is another reason why impositions and obligations may be ineffective. They may be "out of sight, out of mind".

Another example: in China, drivers would never stop for pedestrians at road crossings, or obey other road rules, until cameras were installed and large fines were levied, first in Shenzhen then later in other regions. Suddenly everyone followed the rules as the fines served as a reminder.

Another way to bring people into alignment with rules is to invite them to align, by offering bonuses, benefits, etc. In this version, it becomes a voluntary act. Promise Theory suggests that an initial invitation would foster greater cooperation than imposition.

Example 8 (Referred trust) *Where there is no past memorial experience to rely on, trust can be borrowed or transferred from other currencies, e.g. we trust (or not) someone who speaks well, has rich taste, nice clothes etc. The policy for exchange of these currencies is individual, based on 'memory' (culture, social standing, history, private experience, etc.). We wouldn't normally trust an academic who dressed like a banker (John von Neumann is excused).*

Example 9 (Group referral) *Membership in a group may confer greater trust by borrowing if the assessing agent equates the individual with the group. This is not true transitivity, but individually postulated equivalence.*

Married couples are considered more trustworthy than single persons. Members of elite clubs and institutions, which insist on "standards", may confer a reputational gain in trust. We can trust X people, we can't trust Y people.

Example 10 (Procrastination is non-delivery on a promise) *We don't trust people who are always late, or who don't show up when the say they will with what they've promised. Even if the promise hasn't been made, the expector may infer the existence of a promise as inherited by its understanding of group membership.*

We expect X to be on time as a member of a group we know it is part of. We expect Y to be late as a member of a group we know it belongs to.

This is easily understood as the adjustment to assessing trustworthiness by an agent who expects some outcome b, and therefore has a promise to accept $-b$. The misalignment between the procrastinator in keeping a promise of $+b$ explains the low rated assessment and contempt for the procrastinator. The agent is simply unreliable, as view through the lens of a $-b$ receptor.

Example 11 (Run on the banks) *Trust is the underpinning of the economy. Make a bank insolvent by telling everyone it's insolvent. Silicon Valley Bank, March 2023 one the day of writing this note. The promises made by banks include the ability to retrieve one's money at a promised time.*

Example 12 (Trading Deception risk. Benefit of the doubt) *The simplest expression of trust for cost saving is to trust as a default position. This is not necessarily a moral imperative, it's a lazy evaluation strategy. The opposite of 'benefit of the doubt' is full conspiracy mode*

Axelrod addressed this question of what can be gained by deception. Lying may be used to hack default trust positions. Imitation of species for survival in biology. Wear Gucci handbag, wear suit and tie .. symbols of office by association, we build on referral of trusted symbolics Would you trust a scruffy professor, or one who is dressed like a banker? The promises made by their styles appeal to different audiences. Promise Theory suggests that (whatever they may be) they will tend to align by natural selection.

Example 13 (Secure shell trust question) *The first time one connects to a new server, a fingerprint is presented and the client is asked whether or not the client wants to trust that the identity of the server is as advertised. The user can say no, but then the connection disconnects and not further contact is allowed. If the client says yes, then connection is established and never checked again.*

This is such a short-lived promise (a transaction) that it has little long term significance. It is a typical example of security theatre, in which a check is supposed to make the client feel that the connection will be secure, and the response is to provide a receipt or token of the minimal effort check. In fact the only promise the SSH makes is to encrypt the channel from end to end.

Example 14 (Linux Kernel, pH intrusion prevention, Somayaji) *The pH system does some basic machine learning on n-grams of instructions, and and new unknown sequences are delayed, thus slowing down processes that are not recognized, i.e. not trusted. This adds a cost to the potential attacker, thus pushing the implication of mistrust onto it in the hope that it will go away.*

Example 15 (Websites: from Verisign to Trustpilot) *An example of trust in websites.*

Certificate verification of the website's identity. For online sales, Trustpilot is a site that allows users to write reviews. This has been widely criticized for fake data, with good reviews prioritized and poor reviews suppressed and even removed by the companies being reviewed.

This goes to the point that trust can be undermined by fake or shoddy displays, attempts to blind with science etc.

Example 16 (Algorithms and smart machinery) *Can we trust opaque technologies like chatGPT? If they imitate something we recognize, and therefore hack our promise receptors, they become easier to trust, because agents typically have expectations or receptors for promises they are seeking to align with. Can we trust Google search or Bing? Do we trust the news from BBC, CNN, CCTV, Fox News etc. The arguments around these questions wax and wane over years as different observers gain or lose trust due to occasional anomalies.*

The answers from chatGPT are long and more complex than a search engine's answers, neither of which takes responsibility for the content. Promise Theory suggests that the more complex or specific the output, the less we will tend to trust it.

Example 17 (Explainability) *This is a generalization of the logo argument.*

Communication is the basis for promise sharing. Without communication of events an agent is isolated and trust is unnecessary. Conversely, without trust, communication is ineffective.

The process of explanation (rationalization) of a process, is a resonant relationship between agents, perhaps leading to a key event or decision, is a trust self-building mechanism.

Promise Theory suggests that a simple or general explanation will be more easily trusted than a complex or specific one.

The function of explanation may be to evaluate trust in such ad hoc decisions [29]. The resonance between agents could be internal (virtual agents as propositions and strawmen) in the case of

self-justification, but it could also be an interaction with actual agents in the environment to establish facts. Trust has to be invested in order to learn from the sources and establish trustworthiness.

Any learning process, data gathering process, or chain of dependencies (such as in supply chains and explanatory arguments) involves trust in a sequence of sources and promised inferences. The uncertainty is codified as the intermediate agent theorem in Promise Theory. It follows from the autonomy (causal independence) of each agent. Information without trust is useless, just as with any form of communication. The consequence of the downstream principle is that the ultimate recipient is responsible for believing or trusting in those priors.

Example 18 (Soundbites) *A corollary to the logo and explainability arguments is that people may tend to risk trusting soundbites over detailed explanations, because detailed explanations cost too much to understand. The offer of kinetic trust is a cutoff in the evaluation procedure for the detailed explanation. So when an agent emits soundbites that can be accepted, (all else being equal) the agent will tend to be trusted, unless the agents have a policy to focus on the issue in detail due to particular seriousness.*

Example 19 (Misunderstandings) *When applying for this project the trust dynamics played an interesting role. initially, with no knowledge of NLnet, I was sceptical of the grant system and didn't put too much effort into the application. After being selected, through several rounds, my estimation of the possibility of acceptance rose, but my self-assessed level of trust didn't change significantly. The project was accepted and a call with the moderators explaining the terms led to a wave of optimism and increased positivity towards NLnet, based on the great flexibility promised. I sketched out a plan but this was met with disapproval on several points, expressed as a lack of their trust in the proposed choices, which lowered my trust in response. A failure to reply or respond for several weeks further led to a lowering of trust. Finally, communication was reestablished and the moderators acted quickly to restore trust, answering questions and indicating that their signalling of a lack of trust was now being reversed. Accordingly, my trust rose.*

This example shows how the dynamics of trust are resonant in some sense. That cannot be captured by a simple Prisoner's Dilemma model.

Example 20 (Cold calling) *When sales personnel cold call prospects in the hope of striking up a conversation, they have no history or context.*

When I discover new researchers who excite me, I tend to write to them to express my interest in their work. Perhaps this will lead to future collaboration.

In both cases, such an imposition starts from a potential negative. Promise Theory predicts that this kind of imposition will tend to reduce trust. This seems borne out by experience. Unless the recipient is receptive to a particular need (is aligned with a (-) receptor promise), in which case an imposition will succeed.

So cold calling is a search mechanism for those who are already receptive to an idea. This is generally more successful in sales than in research.

When I receive calls like this myself, I am generally positive to answering them, briefly out of courtesy. That's because I have a personal 'moral policy' to help and support others, if I am in a position to do so. This might be a long term view of selfish behaviour, based on the idea of karma, i.e. being generous of spirit is an investment in future relations, which may be complex and unexpected. But in most cases, cultural conditioning will dominate these interactions, and people will only reply with what they need to for their own benefit, taking a short term view. Tit for tat.

In Japan, cold calling will be summarily ignored.

Example 21 (Brokers, banks, and escrow) *If trust is lacking, one has intermediate agents who can act as Trusted Third Parties (TTP) to broker exchanges.*

Brokers are difficult parties to trust, since they only relay others' promises, which violates the first law (thou shalt not make a promise on behalf of others). The broker typically has little to lose by doing a poor job as jobs are often transactional. Only the reputation may suffer by referrals, but reputation is itself an intermediate brokerage of assessments so trust is ad hoc.

Trust mitigators include the legal system, in which agents expect to be able to sue for damages in case of mistreatment. The escalation of legal texts in the US is an example of mistrust, or the absence of third party regulation. In Norway, by comparison, more regulations are written into law for everyone so that individual contracts can be simplified. On may depend on the provisions of law rather than a expertly written contract subject to ad hoc interpretation by a jury. This is another example of how trust is a cost saving device.

Banks act as trusted third parties in payment exchanges. Escrow services hold money or assets in limbo between agents until reciprocating conditions have been met. This allows someone to take the first step in inviting another to interact, thus building trust, without risking much. In Norway, banks assumed the role of identity providers early in the rise of electronic banking. Even governments rely on BankID, since they need this service themselves and can sell it to others.

Hotels often demand reserve escrow deposits for guests in case of unpaid bills. Money "on the table", placed in temporary suspension or in escrow can serve as a measure of trust for the one making the move, and the request or demand (imposition) is a measure of distrust in the one requesting.

In IT, TTPs are usually only identity validators, which minimally avoid IP spoofing. OpenID (OAuth) as a service is provided by Google, Facebook, and others to provide a simple strong validation of identity, but the third party may mine the data about where identities are being requested for their own profit.

Brokers can go-between agents to preserve their identity or shield them from potential abuse. They can act as impartial agents. In principle, a third party should not have an interest in the outcome of promises they keep for clients, but this is often difficult to argue. For instance, real estate agents make their profit as a fraction of the price of sale, which aligns them more with the seller.

Example 22 (Shared state (information escrow)) *One way to entangle agents in a state of mutual trust is to employ shared state, in which both parties have a state. If both parties in an interaction promise to depend strongly on the information shared by both, this creates a rigid bond between them. This is the nature of entanglement in QM and CS.*

The intermediate agent theorem implies that transitivity of trust cannot be guaranteed. The end to end problem indicates the additional promises needed to overcome these uncertainties. The economics of the centralized broker (manager, leader, etc) are revealed as the cost saving in trusting a single figurehead rather than an N^2 array of trust interactions, each of which is costly to maintain.

Example 23 (The scientific method–process recognition) *This is an example of group association (a norm). Trust in the scientific method as a proxy or trusted third party is an interesting case. If we model this in Promise Theory, one might say that there is a broad consensus (or group) of sources who promise a version of what constitutes a valid method for acquiring knowledge. If there is broad acceptance of this group then a third party agent can observe approximate agreement amongst the practitioners of the method. The method involves building observations based on repeatability. Repetition is the basis of trust building in the observations (as in learning). The trust in sources allows explainability (through theory) to be based on a trusted source. Mathematics is often used as the basis of reasoning since it is repeatable in its steps. What remains to be trusted is the way the path of reasoning is joined together. This is where the downstream principle and intermediate agent theorem apply. This the scientific method acts as a Trusted Third Party guiding extended processes to trustworthiness by constantly calibrating with the method.*

Example 24 (Bureaucracy as group standardization–process recognition) *Standards are calibrating agents.*

The standardization of forms and procedures is a trust building exercise for those who create the system.

The imposition of those standards onto unwilling parties fosters mistrust, as in accordance with the principle that imposition tends to undermine trust.

Example 25 (Deming's theory of profound knowledge) *A set of trust building guidelines for process management, and quality seeking. "Quality" is the assessment of promises (externalized qualities) that are beneficial to the assessor.*

Example 26 (Scandinavian buying habits) *Trust in new thinking is very low. People are very concerned about what others think, and rarely want to stand on their own. Don't want to look stupid? Why is no one else interested in this? The would rather fight over a house than buy it without competition. What's wrong with it?*

Tend to trust the brokers, but they are coin operated.

Example 27 (Oversubscribing seats on a plane) *Airlines who over-book flights present a risk to travellers. The more often the promise of a seat is not honoured, the less trusted the airline.*

Example 28 (Violence against women and minorities) *Will affect their level of trust in society in general. Their assessments leak into other contexts. This alters the default level of trust as a group association. Norms and group prejudices based on identity politics will tend to dominate learned default positions.*

Example 29 (Third party evidence may not convince when trust is undermined) *Lack of trust (fear) blocks rational judgement to the contrary. e.g. mistrust of government during COVID was a symbol of how little actual trust in government remains.*

Example 30 (Adaptive and protectionistic behaviours) *Imagine a cloud system that learns about its clients. It's midnight and I know you are probably drunk, so I'm going to limit your VM purchases to 10, so you don't bring down the East Coast site like you did last year. It's day time, so you probably sobered up, you can have 1000 is you need them now.*

These judgements are conditionals and protectionistic, so we need to be cautious in advocating them. Protectionism doesn't work well, it leads to economic stagnation. Have to give the benefit of the doubt to keep playing else mutual trust is undermined for the long term.

Promise Theory predicts that imposed conditionals will tend to undermine trust.

Example 31 (Shorting stocks) *Shorting a stock means gambling on the price going down.*

"In short selling, a position is opened by borrowing shares of a stock or other asset that the investor believes will decrease in value. The investor then sells these borrowed shares to buyers willing to pay the market price. Before the borrowed shares must be returned, the trader is betting that the price will continue to decline and they can purchase the shares at a lower cost. The risk of loss on a short sale is theoretically unlimited since the price of any asset can climb to infinity."

This feels like betting on mistrust, but that is an incorrect notion of trust. This is not mistrust in the stock, because trust has no negative connotation. It is trust in the reliability of a prediction, or direction of evolution in the market price.

Example 32 (Trust in cryptocurrency) *Initially the promise of BitCoin was independence from banks and regular financial system. Once mainstream promised to accept BitCoin for regular money, it become possible to promise remittances abroad, which are slow and costly in regular national currencies (due to low trust in developing economies). Trust in BitCoin could thus bypass trust through intermediaries. Once BitCoin became tradable, it became simply an asset with a tradable promise of value, at which point it became a speculative virtual asset whose function was something like shares in a public company. It had no remaining functional utility.*

In these phases, users are trading trust in different promises. First it's independence or novelty. Then one overlooks the hackability of the technology to trust in the promise of speed and independence. Then one simply trusts in the market price trend, like in any stock investments.

Notice that in economics, the promise of buying is enough to make the system trade. It's a false assumption that certain goods are valuable because they can be made useful. The usefulness of virtual assets lies only in the extent to which someone else might buy them (for any purpose including no purpose).

Example 33 (Stereotypes, I'm afraid of Americans, foreigners, football fans etc..) *As mentioned, Promise Theory predicts that coarse grained categories like stereotypes will be easy to trust (attractive characterizations), compared to individual assessments. So we may tend to judge people by their feathers rather than as individuals. In other words, we tend to believe in the trustworthiness of the stereotype.*

Isn't this wrong? If we're afraid of these people isn't that the opposite of trusting them more? This is incorrect. We must not confuse trust in the veracity of the stereotype with trust in the behaviour we attribute to the stereotype. It's precisely because we trust our judgement about the use of a stereotype that leads to the unfairness of group classification, when it doesn't represent every individual in the group. Yet we may still tend to trust in the accuracy of the stereotype in spite of knowing better. This is the lazy cost-saving aspect of trust in stereotypes.

9 Thoughts and discussion

The trust literature is huge, but surprisingly contained. There is much redundancy, and few new ideas. The Promise Model goes a long way to explaining the basic dynamics, but researchers continue to try to overcome aspects of trust which they find to be morally confounding. We can do a better job at separating semantics instead of words to deconstruct the scales involved in social currencies. Trust seems to occupy a special role that underpins several close relatives (reliability, confidence, hope, etc).

In Social Sciences, a lot of effort goes into the semantic interpretation of trust, especially in relation to the identities of persons and groups. In Computer Science, trust is treated as a one dimensional quantity: a simple receipt or sometimes probability used to answers a simple 'to be or not to be' question. Initial

levels of trust are thought to be based on the state of an agent before learning about a joint relationship commences (we might call this an intrinsic 'personality' or individual attributes, but it basically amounts to the state of learning up to that moment). This is borne out by neurological studies. Then it is modified by learning, or cognitive state, which we can treat as a local agent's context relative to the processes it has on hand to assess others. The attitude of security experts to trust (as a flaw or weakness) suggests that an *understanding* of trust dynamics is not always what people are really looking for: it's used as a competitive weapon for shaming others. The current obsession with terms like "trust free" or "trustless" systems in IT shows how one crudely attempts to brush trust under a convenient rug. It seems important to clarify these dynamics for a more sustainable future.

Trust involves some common themes. Different trust researchers phrase their observations differently. Paraphrasing the words of the papers reviewed, we can try to summarize the basic issues in a single language aligned with Promise Theory:

- Trust and trustworthiness are different things:

  Trustworthiness is a measure (assessment) of reliability in keeping some promise or collection of promises. It has a compositional aspect when several themes are involved into a common currency or potential.

  Trust given to others (kinetic) is a measure (assessment) of risk appetite or willingness to invest trust in order to save expending one's own resources to achieve an outcome, especially in the face of uncertainty about other agents.

- Trust relies on expectations about some issue. These expectations can be based on previous experiences, reputation, or other factors that influence how much trust we place in someone or something. In PT terms, it suggests receptor promises for the issue to which an agent is sensitive enough to evaluate some level of trust.

- Trust is involved whether to hedge against some risk or potential 'vulnerability'. Some authors overstate the presumed role of vulnerability, or dependency on others. The related concept of "hope" might better capture that neediness. Trust involves a willingness to strategize and gamble on the shape of an interaction. This might involve "moral hazard" (privatization of reward, socialization of risks).

- Trustworthiness increases with reliability, when promises are kept predictably, i.e. integrity. Increased trustworthiness motivates the investment of kinetic trust (risk appetite) from those who assess it.

  Social commentators often suggest this means that the other person or thing is guided by a set of moral principles or values that align with one's own, but this seems to be a speculative assumption. A purely resource based model is simpler and can easily model moral issues too.

- Non-linearity of evolution in time: Trust usually takes a steady accumulation of evidential experience in 'interaction proper time' to build up, but it can quickly be destroyed. It can also be quickly repaired if doubts can be eliminated. This is a prediction of game theory if one treats trust as a kind of payoff or utility.

  There are three concepts of time involved in trust:

  - The internal time of the promiser agent allocated to keep its exterior promise.
  - The internal time of the promisee agent allocated to accepting and assessing trustworthiness of the promiser.
  - The time of a third party observing both agents exterior signals.

- Quasi-mathematical ideas like transitivity of trust or trustworthiness assessments are very often taken as wishful thinking by authors hoping to formalize the ideas. This notion is incompatible with Promise Theory.

- Invitation and transparency often build trust, by offering something freely and unconditionally.

  Offer of a gift on signing up, help in setting up complex requirements can help sweep people off their feet, as long as the gifts are not considered bribes to paper over the cracks.

  When facing complexity, hand holding clients through the perceived risks (User Experience or UX) is a way to win trust. After this initial hand holding, reliability and low touch self-sufficiency is a way to build trust.

Initial relationships support a critical trust threshold, then the cost of the relationship starts to become a diminishing return.

- Trust may be increased by repetition or by memorable impressive displays or dazzling results (peacock display). Showing leadership or taking the initiative. This is another form of invitation vs imposition.
- Trust may be undermined by fake or shoddy displays or behaviours, attempts to bribe, deceive, put on an act, lie, or even blind with science. Incredible advertising can thus work against companies. State sponsored propaganda may be rejected as fake, and further weaken trust if there is already weak trust in government, leading to a downward spiral.

  The learning nature (Bayesian priors) of trust means that these spirals will tend to be up or down and unstable.
- The sociological literature is frequently concerned with the moral dimension to semantics in what contributes to a level of trust. This may not be the right question to ask—indeed moral interpretations are often misleading (as in the stereotype example). More important, we can ask what is the exchange value of process currency. Just as money is more important than any single trade or transaction, or even the reasons for it, the large scale economic picture can be defined in many cases in purely economic and monetary terms. If nothing else, common currencies, are an organizing principle that separate semantics from dynamics, through rules for interaction.
- In order to operationalize trust as a predictive quantity, Burgess suggests that it might be desirable to treat it in the same way as energy [3], using equations that provide the relationship between trust and operational or behavioural measures and the interchange value transfer from the work done. In practice, this is a way of encoding the meanings of trust without relying on vague qualitative descriptions.
- A fairly common theme [54] is to point out that a position of vulnerability by (-) might increase the the hope or level of affinity to trust in a promise (+). e.g. we really need to sell the house and we trust you to offer a good price, or I need a cure so I trust that your medicine will help me. Agents might feel differently if they were less desperate. Again, this is consistent with the idea that decisions are shaped by emotional potentials and only justified by rational means.
- On a simple level, there's something to the observation that "trust vs distrust" amounts to an emotional assessment of "satisfaction vs dissatisfaction" about the assessment of the agent counterpart offering the promise, either in a single episode or over time.
- Neuroscience confirms this emotional basis of trust through hormonal pleasure triggers, and innateness of initial assessments, followed by reinforcement learning, as one would expect from simple functional arguments.
- The cognitive capacity to care about something (i.e. elicit a preference, like an emotional spike) is key to trusting, as noted in the neurological association to pleasure.
- Signalling of trust may be promise or imposition: "You don't trust me! You don't support me!" is a strategy for shaming the (+) party into promising more by the (-) party, but since this is an accusation, which in turn is an imposition, it's more likely to have the opposite effect of undermining trust offered by (+) to (-).
- The key point is that trust (although never mutually assured) is a function of two parties: promises kept (+) and expectations received (-) and, of course, the assessment by the recipient.
- Probability is a seductive method for scientists and engineers, because it seems to solve the problem of quantitative values. In fact, it can be misleading as it covers all its tracks and lacks the dimensionality to be a faithful representation of the original data. Several authors point out these limitations.

Trust is a slowly varying potential with short term fluctuations. More important than small term fluctuations in trust is the long term record of quickness in repairing a defective promise. Repair is the key to restoring trust. If it takes too long, it's gone, but there is a window of redemption associated with some memory scale. This sets a scale from the sampling rate for assessment by the trustee. This is the classic observation of risk management, to repair quickly before side effects can be compounded spiral out of control.

10 Questions

1. How shall we use our model of trust to lubricate human-technology (cyborg) relations in an increasingly augmented semi-virtual world?
2. Could one use trust in friends and acquaintances to form more trustworthy groups and build on that coarse graining to de-risk interactions online with strangers? Just as going with a friend or partner to a human meeting can help.
3. Shall we pay attention to the eigenvalue condition for global referred trust, i.e. reputation? [9]
4. Trust gives authority to agents, and this has resonances, so it's going to be non-linear [18]. Does this help to identify anomalies? What scale should we assume? IT always looks for simple threshold behaviour, but that is a poor guide.
5. There is a tendency to be sceptical of automation. People need mechanisms to feel that they are in control, or that someone else has obviously better control than they do (by qualification or ability). What signals confirm trustworthiness in online interactions, and on what timescale?

   The future of trust management may, after all, be human call centres (and inevitably AI avatars that mimic them). Likely, the proof that one is actually talking to a trained human empathizer will be an important facilitator in trust management.

These questions will be answered in a separate, forthcoming document.

11 Summary

The terms "trust" and "trustworthiness" are used casually in most disciplines. The decision to offer trust and to assess trustworthiness in others is a policy decision, often given with the conviction of ad hoc moral positions. Trust in one's sense of trust even explains why we don't care to define it more carefully.

The goal of these notes has been to extract an impartial core from the literature, in order to understand the role of trust in safety and risk assessments across a wide variety of contexts—to provide a modern view for our emerging socio-technological world. After surveilling the literature, and picking some examples, listed in the references, Promise Theory remains a plausible arbiter of all the cases considered, and offers a clearer picture that can now be written down in a follow up paper.

How do we choose how much trust to assign (how many sampling chances before disengagement) compared to the assessment of trustworthiness? We have to separate assessments of kinetic trust (risk appetite) from potential trust (trustworthiness). The absence of any universality at the level of agents needn't be a hindrance to formalizing these. We might know that someone is unreliable but still choose to tolerate that in a gamble. Consider one last example:

Example 34 (Bargaining or deceiving?) *If a seller advertises a rug for 100 coins, and we assess that it isn't worth that. As we leave, he says 50 coins! We could either bargain for a lower price, trusting that he is basing his price on the assessment that bargaining was his initial promise, or we consider that he was lying and was dishonest. This choice is influenced by the environmental norms, so the interaction is not really one to one. In mixes assessments about the "culture" or context in which the interaction happens (this is Putnam's point). It's ultimately a policy decision for the agents concerned, which is why broad alignment is the key to breadth and depth of the currencies of trust.*

Many authors speak of culture in a case like this. I prefer to speak of an embedded interaction, which is part of a larger process. The process view is deeper and more usefully operational. Culture is itself only the image of ongoing processes with its own memorized norms and emergent patterns.

In summary, using the Promise Theory model, one can distinguish trust and trustworthiness as kinetic and potential forms of behaviour:

- Kinetic trust is the willingness to engage, thus it is rationally related to risk appetite, not to the assessment of risk itself. The decision to offer trust is a policy decision by an agent. It is not a rational decision, but the decision may be rationalized with the help of semantics defined by promises.
- Potential trust is trustworthiness (tendency to keep promises) and is measured by consistency or repeatability of behavioural patterns, i.e. the assessment that an agent keeps its promises. Many semantic embellishments can be added to this basic definition.

- Trust and trustworthiness pertain to specific promises and processes.
- Trust and trustworthiness are assessments made by independent agents (observers) which get adjusted up and down by interactions.
- The most effective strategy for maintaining high assessments of trustworthiness in other agents is to be ready for rapid recovery when incidents that undermine trustworthiness occur (when promises can't be kept).
- The exchange rates for trust between different processes play a role in the impact on each individual agent. Each agent may be viewed as having policies to determine its exchanges rates, just like currency markets. Each agent is essentially a sovereign nation.

  In order for trust to work as as consistent potential, we should measure its changes only in relation to work done. If work is done to validate a promise, then trust kinetic trust is reduced on the ledger of interactions. It make sense to define kinetic trust level is proportional to the sampling rate for assessing 'promise kept'.

  This requires only self-trust in one's assessments. However, agents can also temper others' choices by looking at the behaviours others promise, and try to shape them using incentives that encourage low risk (trustworthy) behaviours. This is now trust can be 'hacked' for manipulation.

Trust management is a crucial topic for the future governance of human-technological society. The future of trust management in the Internet will most likely have to follow the same methods as the real world—because it's for humans to assess. Consider the example of hotels, which hold deposits from clients on check in. This, for instance, would be a way to avoid (D)DoS attacks by asking clients to put "skin in the game" for smaller scale infractions.

Trusted providers of escrow services will be needed here. So far the trust companies have not been wholly convincing. The online payment companies, for example, have been remiss in providing easily available services, or are too greedy with fees. The question remains who will pay for trust management services. If a service were offered in a clever way, the providers would benefit from the trust built in themselves as trust providers. This in turn could attract business for other transaction fees to cover the costs. Telecommunications providers currently benefit from SMS charges in two-factor authentications, and other identity transactions. This is another route by which costs can be covered. We have to expect that, in the future, more payments will pass through personal electronic devices. Autonomy is unlikely to go into reverse.

No direct examples have come up where the empirical evidence or the philosophy of trust disagrees in spirit with hypotheses of Promise Theory, but some phrasings need clarification on both sides to make assumptions clear. This suggests that the promise model is a good avenue for future development.

Promise Theory suggests that:

- Impositions tend to misalign trust (cold calling, top down rules, etc).
- Unconditional invitation is a protocol to win trust, e.g. money on the table, escrow, I'll go first. But conditional invitations are often impositions, threats, e.g. if you don't give me X you can't do Y.
- A simple or general explanation will be more easily trusted than a complex or specific one (scaling).
- Repetition (rote) is a way to hack trust.
- Group membership may be used to infer trust by appealing to a coarse group promise (inheritance).
- Agents' finite resources imply limited time or activity to verify promises. Trust is a waiving of verification and therefore a resource substitute. Shared finite resources or state tends to create conflicts of interest, which trust seeks to eliminate (tragedy of commons).

The next step is to write down the consistent resource model for trust in Promise Theory, which is compatible with all the foregoing observations.

Acknowledgment: This work is supported by NLnet project Trust Semantic Learning and Monitoring. I'm grateful to Edmund Humenberger for helpful discussions.

References

[1] W. Stone. Measuring social capital: Towards a theoretically informed measurement framework for researching social capital in family and community life. Technical Report research paper 24, Australian Institute of Family Studies, 2001.

[2] H. and K. Yasunobu. What is social capital? a comprehensive review of the concept. *Asian Journal of Social Science*, 37(3):480–510, 2009.

[3] M. Burgess. Notes on trust as a causal basis for social science, v0.2. *SSRN Archive, available at http://dx.doi.org/10.2139/ssrn.4252501*, August 2022.

[4] D. Gambetta. *Trust: Making and Breaking Cooperative Relations*, chapter Can We Trust Trust?, pages 213–237. Blackwell, 08 2000.

[5] J.F. Nash. *Essays on Game Theory*. Edward Elgar, Cheltenham, 1996.

[6] R.B. Myerson. *Game theory: Analysis of Conflict*. (Harvard University Press, Cambridge, MA), 1991.

[7] R. Axelrod. *The Complexity of Cooperation: Agent-based Models of Competition and Collaboration*. Princeton Studies in Complexity, Princeton, 1997.

[8] R. Axelrod. *The Evolution of Co-operation*. Penguin Books, 1990 (1984).

[9] J.A. Bergstra and M. Burgess. Local and global trust based on the concept of promises. Technical report, arXiv.org/abs/0912.4637 [cs.MA], 2006.

[10] J.A. Bergstra and M. Burgess. *Promise Theory: Principles and Applications (second edition)*. $\chi t Axis$ Press, 2014,2019.

[11] E. Brousseau and J-M. Glachant, editors. *The Economics of Contracts Theory and Applications*. Cambridge University Press, 2002.

[12] B. Salani'e. *The Economics of Contracts (second edition)*. MIT Press, 2005.

[13] E. Ostrom. *Governing the Commons*. Cambridge, 1990.

[14] E. Ostrom. *Understanding Institutional Diversity*. Princeton University Press, 2005.

[15] R. Dunbar. *Grooming, Gossip and the Evolution of Language*. Faber and Faber, London, 1996.

[16] W.X. Zhou, S. Sornette, R.A. Hill, and R.I.M. Dunbar. Discrete hierarchical organization of social group sizes. *Proc. Royal Soc.*, 272:439–444, 2004.

[17] K. Farrahi and K. Zia. Trust reality-mining: evidencing the role of friendship for trust diffusion. *Hum. Cent. Comput. Inf. Sci.*, 2017.

[18] M. Burgess. Authority (i): A promise theoretic formalization. *SSRN: https://ssrn.com/abstract=3855352*, http://dx.doi.org/10.2139/ssrn.3855352, 2021.

[19] S. Manigart, M. Korsgaard, R. Folger, H. Sapienza, and K. Baeyens. The impact of trust on private equity contracts. 03 2002.

[20] P.H. Kim, C.D. Cooper, K.T. Dirks, and D.L. Ferrin. Repairing trust with individuals vs. groups. *Organizational Behavior and Human Decision Processes*, 120:1–14, 2013.

[21] S.A. Rands and C.C. Ioannou. Personality variation is eroded by simple social behaviours in collective foragers. *PLOS Computational Biology*, 19(3):1–22, 03 2023.

[22] M. Kumove. Rent-free in your head? how generalised trust is affected by the trust and salience of outgroups. *Social Indicators Research*, pages 1–26, 02 2023.

[23] R. Putnam. *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press, 1993.

[24] F. Fukuyama. *Trust*. New York Free Press, 1995.

[25] A. Valadbigi and B. Haratyunyan. Trust: The social virtues and the creation of prosperity by francis fukuyama (review). *Studies of Changing Societies: Comparative and Interdisciplinary Focus*, 1(1):80–95, 2012.

[26] R.C. Mayer, J.H. Davis, and F.D. Schoorman. An integrative model of organizational trust. *The Academy of Management Review*, 20(3):709–734, 1995.

[27] E.L. Glaeser, D.I. Laibson, J.A. Scheinkman, and C.L. Soutter. Measuring Trust*. *The Quarterly Journal of Economics*, 115(3):811–846, 08 2000.

[28] J. Bergstra and M. Burgess. *Money, Ownership, and Agency.* χt-axis Press, 2019.

[29] H. Mercier and D. Sperber. *The Enigma of Reason, A New Theory of Human Understanding.* Penguin, 2017.

[30] B. Robbins. Measuring generalized trust: Two new approaches. *Sociological Methods & Research*, 51:305–356, 02 2021.

[31] J. Brahm and W. Rahn. Individual-level evidence for the causes and consequences of social capital. *American Journal of Political Science*, 41:999–1023, 1997.

[32] P. Paxton. Is social capital declining in the united states? a multiple indicator assessment. *American Journal of Sociology*, 105:88–127, 1999.

[33] K. Newton and S. Zmerli. Three forms of trust and their association. *European Political Science Review*, 3:169–200, 2011.

[34] M. Freitag and Richard Traunmüller. Spheres of trust: An empirical analysis of the foundations of particularized and generalized trust. *European Journal of Political Research*, 48:782–803, 2009.

[35] R. Axelrod. An evolutionary approach to norms. *American Political Science Review*, 80(4):1095–1111, 1986.

[36] R. Axelrod. *Genetic Algorithms and Simulated Annealing (ed L. Davis)*, chapter The Evolution of Strategies in the Iterated Prisoner's Dilemma, pages 32–41. Pittman, 1987.

[37] L.M.A. Bettencourt. The origins of scaling in cities (with supplements). *Science*, 340:1438–1441, 2013.

[38] L.M.A. Bettencourt, J. Lobo, D. Helbing, C. Hühnert, and G.B. West. Growth, innovation, scaling and the pace of life in cities. *Proceedings of the National Academy of Sciences*, 104(107):7301–7306, 2007.

[39] Restoring trust in financial markets. OECD website.

[40] OECD. Chapter 1. trust and financial markets. In *OECD Business and Finance Outlook 2019 : Strengthening Trust in Business.*

[41] S. Huck, G.K. Lünser, and J-R. Tyran. Pricing and trust. CEPR Discussion Paper No. DP6135, Available at SSRN: https://ssrn.com/abstract=1133780, February 2007.

[42] Alvin J. Huss Professor of Management and Kellogg School of Management Strategy. Trust in transactions: An economist's perspective. The Trust Project at Northwestern University/Videos.

[43] Victoria L. Lemieux. *From Trustless Trust to "The Great Chain of Certainty"*, page 4970. Cambridge University Press, 2022.

[44] P. De Filippi, M. Mannan, W. Reijers, P. Berman, and J. Henderson. Blockchain technology, trust & confidence: Reinterpreting trust in a trustless system? *SSRN available at https://ssrn.com/abstract=4300486 or http://dx.doi.org/10.2139/ssrn.4300486*, page 21, December 12th 2022.

[45] M. Morvan and S. Sené. A Distributed Trust Diffusion Protocol for Ad Hoc Networks. In *Second International Conference on Wireless and Mobile Communications*, page 87, Bucarest, Romania, 2006.

[46] M.A. Orangi and A. Hashemi Golpayegani. An activity-based user trusting behavior diffusion model in social networks. In *2018 9th International Symposium on Telecommunications (IST)*, pages 32–38, 2018.

[47] J.A. Bergstra and M. Düwell. Accusation theory. *Transmathematica*, Dec. 2021.

[48] A. Jøsang. Trust and reputation systems. *Lecture Notes on Computer Science*, 4677, 2007.

[49] M. Graydon and C. Holloway. An investigation of proposed techniques for quantifying confidence in assurance arguments, 05 2016.

[50] Z. Noorian and M. Ulieru. The state of the art in trust and reputation systems: A framework for comparison. *Journal of Theoretical and Applied Electronic Commerce Research*, 5(2):97–117, 2010.

[51] A. Wong, X.Y. Wang, and A. Hryniowski. How much can we really trust you? towards simple, interpretable trust quantification metrics for deep neural networks, 2020.

[52] A. Hryniowski, X.Y. Wang, and A. Wong. Where does trust break down? a quantitative trust analysis of deep neural networks via trust matrix and conditional trust densities, 2020.

[53] M. Ruby. How chatgpt works: The model behind the bot. Towards Data Science (Medium), 2023 January.

[54] O. Vereschak, G. Bailly, and B. Caramiaux. How to evaluate trust in ai-assisted decision making? a survey of empirical methodologies. *Proc. ACM Hum.-Comput. Interact.*, 5(CSCW2), oct 2021.

[55] A. Shamekhi amd Q.V. Liao amd D. Wang, Rachel.K.E. Bellamy, and T. Erickson. Face value? exploring the effects of embodiment for a group facilitation agent. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI 18). ACM*, 2018.

[56] F. and A. Meyer-Lindenberg. Toward a model of interpersonal trust drawn from neuroscience, psychology, and economics. *Trends in Neurosciences*, 42(2):92–101, 2019.

[57] P.J. Zak. The neuroscience of trust: Management behaviors that foster employee engagement. *Havard Business Review*, 2017.

[58] Steven L. Bressler and Vinod Menon. Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, 14(6):277–290, 2010.

[59] C. Clark. Trust diffusion: The effect of interpersonal trust on structure, function, and organizational transparency. *Business & Society - BUS SOC*, 44:357–368, 09 2005.

[60] M. Fujii. Simulations of the diffusion of innovation by trustdistrust model focusing on the network structure. *Rev Socionetwork Strat*, 16:527–544, 2022.

[61] D. Mezick and M. Sheffield. *Inviting Leadership: Invitation-based Change*. Freestanding Press, 2018.

[62] E.H. Klijn, J. Edelenbos, , and B. Steijn. Trust in governance networks: Its impacts on outcomes. *Administration & Society - ADMIN SOC*, 42:193–221, 04 2010.

[63] X. Zhang, H. Wang, J. Nan, Y. Luo, and Y. Yi. Modeling and numerical methods of supply chain trust network with the complex network. *Symmetry*, 14:235, 01 2022.

[64] S.E. Fawcett, G.M. Magnan, and A.J. Williams. Supply chain trust is within your grasp. *Supply Chain Management Review*, 8:20–26, 01 2004.